Quantifying identifiability in independent component analysis

Sokol, Alexander; Maathuis, Marloes H.; Falkeborg, Benjamin

Published in:
Electronic Journal of Statistics

DOI:
10.1214/14-EJS932

Publication date:
2014

Electronic Journal of Statistics
Vol. 8 (2014) 1438–1459
ISSN: 1935-7524
DOI: 10.1214/14-EJS932

Quantifying identifiability in independent component analysis

Alexander Sokol*

Institute of Mathematics
University of Copenhagen
2100 Copenhagen, Denmark
e-mail: alexander@math.ku.dk

Marloes H. Maathuis

Seminar für Statistik
ETH Zürich
8092 Zürich, Switzerland
e-mail: maathuis@stat.math.ethz.ch

Benjamin Falkeborg

Department of Economics
University of Copenhagen
2100 Copenhagen, Denmark
e-mail: benjamin.falkeborg@econ.ku.dk

Abstract: We are interested in consistent estimation of the mixing matrix in the ICA model, when the error distribution is close to (but different from) Gaussian. In particular, we consider n independent samples from the ICA model $X = A\epsilon$, where we assume that the coordinates of ϵ are independent and identically distributed according to a contaminated Gaussian distribution, and the amount of contamination is allowed to depend on n. We then investigate how the ability to consistently estimate the mixing matrix depends on the amount of contamination. Our results suggest that in an asymptotic sense, if the amount of contamination decreases at rate $1/\sqrt{n}$ or faster, then the mixing matrix is only identifiable up to transpose products. These results also have implications for causal inference from linear structural equation models with near-Gaussian additive noise.

MSC 2010 subject classifications: Primary 62F12; secondary 62F35.
Keywords and phrases: Independent Component Analysis, LiNGAM, Identifiability, Kolmogorov norm, Contaminated distribution, Asymptotic statistics, Empirical process, Linear structural equation model.

1. Introduction

We consider the p-dimensional independent component analysis (ICA) model

$$X = A\epsilon, \tag{1.1}$$

where A is a $p \times p$ mixing matrix, ϵ is a p-dimensional error (or source) variable with independent and nondegenerate coordinates of mean zero, and X is a p-dimensional observational variable. Based on observations of X, ICA aims to

estimate the mixing matrix A and the distribution of the error variable ϵ. Theory and algorithms for ICA can be found in, e.g., [4, 5, 11, 12, 13, 18, 22]. ICA has applications in many different disciplines, including blind source separation (e.g., [6]), face recognition (e.g., [2]), medical imaging (e.g., [3, 15, 26]) and causal discovery using the LiNGAM method (e.g., [23, 24]).

Our focus is on consistent estimation of the mixing matrix. Here, identifiability is an issue, since two different mixing matrices A and B may yield the same distribution of X. This is the case, for example, if the distribution of ϵ is multivariate Gaussian and $AA^t = BB^t$. In this case, the mixing matrix cannot be identified from X, instead only the transpose product of the mixing matrix can be identified. In [5], it was shown that whenever at most one of the components of ϵ is Gaussian, the mixing matrix is identifiable up to scaling and permutation of columns. This result was expanded upon in the ICA context in Theorem 4 of [9] and extended considerably to the broader class of additive index models in Theorem 1 of [27]. In order to illustrate the relevance of the mixing matrix in (1.1), we give an example based on causal inference.

Example 1.1. Consider a two-dimensional linear structural equation model with additive noise of the form

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}, \tag{1.2}$$

see, e.g., [23]. We assume that the coordinates of ϵ are independent, nondegenerate and have mean zero, and are independent of X. We also assume that C is strictly triangular, meaning that all entries of C are zero except either C_{12} or C_{21}. In the context of linear structural equation models, identifying C corresponds to identifying whether X_1 is a cause of X_2 (corresponding to $C_{21} \neq 0$), X_2 is a cause of X_1 (corresponding to $C_{12} \neq 0$), or neither is a cause of the other (corresponding to $C_{21} = C_{12} = 0$).

As C is strictly triangular, $I - C$ is invertible. Letting $A = (I - C)^{-1}$, we obtain

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}, \tag{1.3}$$

where A is upper or lower triangular according to whether the same holds for C. Thus, we have arrived at an ICA model of the form (1.1). In the case where ϵ is jointly Gaussian, it is immediate that we cannot identify A from the distribution of X alone. By the results of [5, 23], identification of A up to scaling and permutation of columns from the distribution of X is possible when ϵ has at most one Gaussian coordinate. In this case, we may therefore infer causal relationships from estimation of the mixing matrix in an ICA model. The LiNGAM method [23] is based on this idea. However, if we only get n i.i.d. samples from X and the distribution of ϵ is close to Gaussian, it may be expected that estimation of A becomes difficult, and consequently causal inference is difficult as well. ∘

Motivated by the above, we take an interest in the following question: When the distribution of ϵ is close to Gaussian but non-Gaussian, how difficult is it

to consistently estimate the mixing matrix as the number n of observations increases? For any fixed non-Gaussian ϵ, the results of [5] show that A is identifiable up to scaling and permutation of columns, so letting n tend to infinity in this model, we should be able to consistently estimate A up to such scaling and permutation of columns. However, if the distribution of ϵ converges to a Gaussian distribution, it is natural to expect that the model begins to take on the properties of ICA models with Gaussian errors, where only the transpose product AA^t is identifiable. Our question, heuristically speaking, is: How large should the number of observations n be in order to counterbalance the near-Gaussianity of the noise distribution? We think of this as quantifying the level of identifiability of A. In order to elucidate this issue, we will consider asymptotic scenarios where the distribution of ϵ depends on the sample size n, and tends to a Gaussian distribution as n tends to infinity.

2. Problem statement and main results

ICA can be used to estimate A when the distribution of ε is unknown. In this case, we may think of the statistical model (1.1) as the collection of probability measures

$$\{L_A(R) \mid A \in \mathbb{M}(p,p), R \in \mathcal{P}(p)\}, \tag{2.1}$$

where $\mathbb{M}(p,p)$ denotes the space of $p \times p$ matrices, $L_A : \mathbb{R}^p \to \mathbb{R}^p$ is given by $L_A(x) = Ax$, $L_A(R)$ denotes the image measure of R under the transformation L_A, $\mathcal{P}(p)$ denotes the set of product probability measures on $(\mathbb{R}^p, \mathcal{B}_p)$ with nondegenerate mean zero coordinates and $\mathcal{B}_p$ denotes the Borel σ-algebra on $\mathbb{R}^p$. With ε having distribution R, this means that the error distribution has independent nondegenerate mean zero coordinates. In other words, it is assumed that the distribution of X in (1.1) is equal to $L_A(R)$ for some $A \in \mathbb{M}(p,p)$ and $R \in \mathcal{P}(p)$. This is a semiparametric model, where A is the parameter of interest and R is a nuisance parameter. Asymptotic distributions of estimates of the mixing matrix in this type of set-up are derived in, e.g., [1, 4, 14]. The difficulty of estimating A can then be appraised by considering for example the asymptotic variance of the estimates.

Alternatively, one can consider estimation of A for a given error distribution. This is the approach we take in this paper. When ε has the distribution of some fixed $R \in \mathcal{P}(p)$, the statistical model (1.1) is the collection of probability measures

$$\{L_A(R) \mid A \in \mathbb{M}(p,p)\}. \tag{2.2}$$

Results on identifiability of A in (2.2) follow from the results of [5] and [9]. In particular, if no two coordinates of R are jointly Gaussian, the mixing matrix A is identifiable up to sign reversion and permutation of columns, in the sense that $L_A(R) = L_B(R)$ implies $A = B\Lambda P$ for some diagonal matrix Λ with $\Lambda^2 = I$ and some permutation matrix P.

We are interested in how difficult it is to consistently estimate the mixing matrix in (2.2) when the error distributions are different from Gaussian but close to Gaussian. Some results in this direction can be found in [19], where the authors calculated the Crámer-Rao lower bound for the model (2.2), under the assumption that the coordinates of the error distribution have certain regularity criteria such as finite variance and differentiable Lebesgue densities. These results indicate how the minimum variance of an unbiased estimator of the mixing matrix depends on the error distribution.

We consider the problem from the following different perspective. For $p \geq 1$ and any signed measure μ on $(\mathbb{R}, \mathcal{B})$, let $\mu \otimes \mu$ denote the product measure of μ with itself, and let $\mu^{\otimes p} = \otimes_{i=1}^{p} \mu$ denote the p-fold product measure. Fix two nondegenerate mean zero probability measures ξ and ζ with $\xi \neq \zeta$, and let $P_e(\beta)$ be the contaminated distribution given by

$$P_e(\beta) = \beta\xi + (1-\beta)\zeta. \tag{2.3}$$

We write F_β^A for the cumulative distribution function of $L_A(P_e(\beta)^{\otimes p})$, and write $F^A = F_0^A$. Note that F^A is then the cumulative distribution function of $L_A(\zeta^{\otimes p})$. We assume that we observe n i.i.d. observations from the distribution $F_{\beta_n}^A$, where the amount of contamination β_n is allowed to depend on the sample size. Our results indicate that, in this framework, consistent estimation of A (up to scaling and permutation of columns) cannot be expected when $\beta_n = o(1/\sqrt{n})$ and ζ is mean zero Gaussian.

3. An upper asymptotic distance bound

In this section, we develop some preliminary results which will be used to prove our main results in Section 4. We begin by introducing some notation. For any measure μ on $(\mathbb{R}^p, \mathcal{B}_p)$, let $|\mu|$ denote the total variation measure of μ, see, e.g., [21]. We define two norms by

$$\|\mu\|_\infty = \sup_{x \in \mathbb{R}^p} |\mu((-\infty, x_1] \times \cdots \times (-\infty, x_p])|, \tag{3.1}$$

$$\|\mu\|_{tv} = |\mu|(\mathbb{R}^p), \tag{3.2}$$

and refer to these as the uniform and the total variation norms, respectively. The uniform norm for measures is also known as the Kolmogorov norm. Note that if P and Q are two probability measures on $(\mathbb{R}^p, \mathcal{B}_p)$ with cumulative distribution functions F and G, then $\|P - Q\|_\infty = \|F - G\|_\infty$. Finally, we use the notation $f(s) \sim g(s)$ as $s \to s_0$ when $\lim_{s \to s_0} f(s)/g(s) = 1$.

As in the previous section, let ξ and ζ be two nondegenerate mean zero probability distributions on $(\mathbb{R}, \mathcal{B})$ with $\xi \neq \zeta$. We aim to bound the distance

$$\|F_\beta^A - F_\beta^B\|_\infty = \|L_A(P_e(\beta)^{\otimes p}) - L_B(P_e(\beta)^{\otimes p})\|_\infty \tag{3.3}$$

for matrices $A, B \in \mathbb{M}(p,p)$ with $F^A = F^B$. To this end, define

$$\nu = (\xi - \zeta)/\|\xi - \zeta\|_\infty. \tag{3.4}$$

The following theorem is a first step towards our goal.

Theorem 3.1. *Let $\beta \in (0,1)$, and let $A \in \mathbb{M}(p,p)$. Then*

$$\lim_{\beta \to 0} \frac{L_A(P_e(\beta)^{\otimes p}) - L_A(\zeta^{\otimes p})}{\|P_e(\beta) - \zeta\|_\infty} = \sum_{k=1}^{p} L_A(\zeta^{\otimes(k-1)} \otimes \nu \otimes \zeta^{\otimes(p-k)}), \tag{3.5}$$

where convergence is in $\|\cdot\|_\infty$. Moreover, F^A_β tends uniformly to F^A as β tends to zero.

The proof of Theorem 3.1 exploits properties of the contaminated distributions $P_e(\beta)$ for $\beta \in (0,1)$, in particular the fact that $\|P_e(\beta)-\zeta\|_\infty$ is nonzero and linear in β and that $(P_e(\beta)-\zeta)/\|P_e(\beta)-\zeta\|_\infty$ is constant in β. These properties are used to obtain a decomposition of $P_e(\beta)^{\otimes p}$ as a polynomial function of β in the proof. As Lemma 3.2 shows, only contaminated distributions have these properties. This is our main reason for working with this family of distributions.

Lemma 3.2. *Let $\beta \mapsto Q(\beta)$ be a mapping from $(0,1)$ to the space of probability measures on $(\mathbb{R}, \mathcal{B})$ with the properties that $\|Q(\beta) - \zeta\|_\infty$ is nonzero and linear in β and $(Q(\beta) - \zeta)/\|Q(\beta) - \zeta\|_\infty$ is constant in β. Then $Q(\beta)$ can be written as a contaminated ζ distribution, in the sense that $Q(\beta) = \beta\xi + (1-\beta)\zeta$ for some probability measure ξ on $(\mathbb{R}, \mathcal{B})$.*

Due to the properties of contaminated distributions, Theorem 3.1 in fact also holds for other norms than the uniform norm. However, the choice of the norm is important when we wish to bound the norm of the right-hand side of (3.5). Such a bound is achieved in Lemma 3.3.

Lemma 3.3. *Let $A \in \mathbb{M}(p,p)$. Then*

$$\left\| \sum_{k=1}^{p} L_A \left(\zeta^{\otimes(k-1)} \otimes \nu \otimes \zeta^{\otimes(p-k)} \right) \right\|_\infty \leq 2p. \tag{3.6}$$

Combining Theorem 3.1 and Lemma 3.3 yields the following corollary, which we give without proof.

Corollary 3.4. *Let $A, B \in \mathbb{M}(p,p)$ be such that $F^A = F^B$. Define*

$$\begin{aligned} \Gamma(A,B,\nu) = &\sum_{k=1}^{p} L_A \left(\zeta^{\otimes(k-1)} \otimes \nu \otimes \zeta^{\otimes(p-k)} \right) \\ &- \sum_{k=1}^{p} L_B \left(\zeta^{\otimes(k-1)} \otimes \nu \otimes \zeta^{\otimes(p-k)} \right). \end{aligned} \tag{3.7}$$

Then we have, for $\beta \to 0$,

$$\|F^A_\beta - F^B_\beta\|_\infty \sim \left\| \Gamma\left(A, B, \frac{\xi - \zeta}{\|\xi - \zeta\|_\infty}\right) \right\|_\infty \|P_e(\beta) - \zeta\|_\infty \leq 4p\beta\|\xi - \zeta\|_\infty. \tag{3.8}$$

Corollary 3.4 shows that, if $F^A = F^B$, the distance between the observational distributions F^A_β and F^B_β decreases asymptotically linearly in β as β tends to zero.

The corollary is stated under the condition that $F^A = F^B$. For later use, we characterize the occurrence of this in the next lemma, in terms of ζ, A and B, for the case where A and B are invertible. Recall that a probability distribution Q on $(\mathbb{R}, \mathcal{B})$ is said to be symmetric if, for every random variable Y with distribution Q, Y and $-Y$ have the same distribution. The proof of Lemma 3.5 is a simple consequence of Theorem 4 of [9].

Lemma 3.5. *Let $A, B \in \mathbb{M}(p,p)$ be invertible. Then the following hold:*

1. *If ζ is Gaussian, then $F^A = F^B$ if and only if $AA^t = BB^t$.*
2. *If ζ is non-Gaussian and symmetric, then $F^A = F^B$ if and only if we have $A = B\Lambda P$ for some permutation matrix P and a diagonal matrix Λ satisfying $\Lambda^2 = I$.*
3. *If ζ is non-symmetric, then $F^A = F^B$ if and only if $A = BP$ for some permutation matrix P.*

4. Asymptotic properties of ICA models with near-Gaussian noise

We now use the results obtained in Section 3 to obtain asymptotic results on ICA models with near-Gaussian noise. We will consider a sequence of ICA models with increasingly near-Gaussian noise, and will investigate the asymptotic properties of this sequence of models.

We need some basic facts about random fields in order to formulate our results, see [16] and [17] for an overview. Recall that a mapping $R : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ is said to be symmetric if $R(x,y) = R(y,x)$ for all $x, y \in \mathbb{R}^p$, and is said to be positive semidefinite if for all $n \geq 1$ and for all $x_1, \ldots, x_n \in \mathbb{R}^p$ and $\xi_1, \ldots, \xi_n \in \mathbb{R}$, it holds that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \xi_i R(x_i, x_j) \xi_j \geq 0. \tag{4.1}$$

For any symmetric and positive semidefinite function $R : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$, there exists a mean zero Gaussian random field W with covariance function R and with sample paths in $\mathbb{R}^{\mathbb{R}^p}$. In general, W will not have continuous paths. For a general random field W, we associate with W its intrinsic pseudometric ρ on $\mathbb{R}^p$, given by

$$\rho(x,y) = \sqrt{E(W(x) - W(y))^2}. \tag{4.2}$$

If the metric space $(\mathbb{R}^p, \rho)$ is separable, we say that W is separable. In this case, $\|W\|_\infty = \sup_{x \in D} |W(x)|$ with probability one, for any countable subset D of $\mathbb{R}^p$ which is dense under the pseudometric ρ. In particular, whenever the σ-algebra on the space where W is defined is complete, $\|W\|_\infty$ is measurable.

The following lemma describes some important properties of a class of Gaussian fields particularly relevant to us. The result is well known, see for example [8], and can be proven quite directly using the approximation results in [7].

Lemma 4.1. *Let F be a cumulative distribution function on $\mathbb{R}^p$. There exists a p-dimensional separable mean zero Gaussian field W which has covariance function $R : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ given by $R(x,y) = F(x \wedge y) - F(x)F(y)$ for $x, y \in \mathbb{R}^p$, where $x \wedge y$ is the coordinate wise minimum of x and y. With $\mathbb{Q}$ denoting the rationals, it holds that $\|W\|_\infty = \sup_{x \in \mathbb{Q}^p} |W(x)|$ and $\|W\|_\infty$ is almost surely finite.*

For a fixed cumulative distribution function F, we refer to the Gaussian field described in Lemma 4.1 as an F-Gaussian field. We are now ready to formulate our results on asymptotic scenarios in ICA models.

Theorem 4.2 describes the classical asymptotic scenario, where the error distribution does not depend on the sample size n. Fix a nondegenerate mean zero probability distribution ζ on $(\mathbb{R}, \mathcal{B})$ and a matrix $A \in \mathbb{M}(p,p)$. As in the previous section, we let F^A denote the cumulative distribution function of $L_A(\zeta^{\otimes p})$, corresponding to the distribution of $A\epsilon$ when ϵ is a p-dimensional variable with independent coordinates having distribution ζ. Consider a probability space $(\Omega, \mathcal{F}, P)$ endowed with independent variables $(X_k)_{k \geq 1}$ with cumulative distribution function F^A. Let $\mathbb{F}_n^A$ be the empirical distribution function of $X_1, \ldots, X_n$. Also assume that we are given an F^A-Gaussian field W on $(\Omega, \mathcal{F}, P)$.

Theorem 4.2. *Let $c \geq 0$ be a continuity point of the distribution of $\|W\|_\infty$. Then*

$$\lim_{n \to \infty} P(\sqrt{n}\|\mathbb{F}_n^A - F^A\|_\infty > c) = P(\|W\|_\infty > c), \tag{4.3}$$

while in the case where $F^A \neq F^B$, it holds that

$$\lim_{n \to \infty} P(\sqrt{n}\|\mathbb{F}_n^A - F^B\|_\infty > c) = 1. \tag{4.4}$$

Equations (4.3) and (4.4) roughly state that in the classical asymptotic scenario, $\sqrt{n}\|\mathbb{F}_n^A - F^A\|_\infty$ converges in distribution to $\|W\|_\infty$, while $\sqrt{n}\|\mathbb{F}_n^A - F^B\|_\infty$ is not bounded in probability if $F^A \neq F^B$. Note that Lemma 3.5 gives us conditions for $F^A = F^B$ and $F^A \neq F^B$ depending on ζ.

Next, we consider an asymptotic scenario where the error distribution is contaminated and the amount of contamination depends on the sample size n. As in Section 3, ξ and ζ are fixed nondegenerate mean zero probability measures on $(\mathbb{R}, \mathcal{B})$ with $\xi \neq \zeta$, $P_e(\beta) = \beta\xi + (1-\beta)\zeta$, $A \in \mathbb{M}(p,p)$ is a fixed matrix, F^A is the cumulative distribution function of $L_A(\zeta^{\otimes p})$ and F_β^A is the cumulative distribution function of $L_A(P_e(\beta)^{\otimes p})$. Thus, F_β^A is the cumulative distribution function of $A\varepsilon$, where ε is a p-dimensional variable with independent coordinates having distribution $P_e(\beta)$. Consider a sequence (β_n) in $(0,1)$, and consider a probability space $(\Omega, \mathcal{F}, P)$ endowed with a triangular array $(X_{nk})_{1 \leq k \leq n}$ such that for each n, the variables $X_{n1}, \ldots, X_{nn}$ are independent variables with cumulative distribution function $F_{\beta_n}^A$. Let $\mathbb{F}_{\beta_n}^A$ be the empirical distribution function

of $X_{n1}, \ldots, X_{nn}$. Also assume that we are given an F^A-Gaussian field W on $(\Omega, \mathcal{F}, P)$. We are interested in the asymptotic properties of $\mathbb{F}^A_{\beta_n}$. Theorem 4.3 is our main result for this type of asymptotic scenarios.

Theorem 4.3. *Let* $\lim_n \sqrt{n}\beta_n = k$ *for some* $k \geq 0$. *If* $F^A = F^B$, *then*

$$\begin{aligned} P(\|W\|_\infty > c + 4pk\|\xi - \zeta\|_\infty) &\leq \liminf_{n\to\infty} P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) \\ &\leq \limsup_{n\to\infty} P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) \\ &\leq P(\|W\|_\infty \geq c - 4pk\|\xi - \zeta\|_\infty). \end{aligned} \tag{4.5}$$

In particular, if $k = 0$ *and* c *is a continuity point of the distribution of* $\|W\|_\infty$, *we have*

$$\lim_{n\to\infty} P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) = P(\|W\|_\infty > c). \tag{4.6}$$

Theorem 4.3 essentially shows that for the asymptotic scenario considered, the convergence of $F^A_{\beta_n}$ to F^A is fast enough to ensure that the asymptotic properties of $\mathbb{F}^A_{\beta_n}$ are determined by F^A instead of $F^A_{\beta_n}$. Corollary 4.4 applies this result to the case where the error distributions become close to Gaussian without being Gaussian.

Corollary 4.4. *Assume that* $\lim_n \sqrt{n}\beta_n = 0$. *Let* $A, B \in \mathbb{M}(p,p)$ *be invertible. Assume that* $AA^t = BB^t$ *while* $A \neq B\Lambda P$ *for all diagonal* Λ *with* $\Lambda^2 = I$ *and all permutation matrices* P. *Let* ζ *be a nondegenerate Gaussian distribution and let* ξ *be such that* $P_e(\beta)$ *is non-Gaussian for all* $\beta \in (0,1)$. *Let* c *be a point of continuity for the distribution of* $\|W\|_\infty$, *with* W *an* F^A*-Gaussian field. Then*

1. $F^A_{\beta_n} \neq F^B_{\beta_n}$ *for all* $n \geq 1$.
2. $\lim_{n\to\infty} P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) = P(\|W\|_\infty > c)$.

Statement (1) of Corollary 4.4 shows that for any finite n, we are in the case where, were the error distribution not changing with n, it would be possible to asymptotically distinguish $F^A_{\beta_n}$ and $F^B_{\beta_n}$ at rate $1/\sqrt{n}$ as in (4.4) of the classical case. However, statement (2) shows that as n increases and the error distribution becomes closer to a Gaussian distribution, distinguishing $F^A_{\beta_n}$ and $F^B_{\beta_n}$ at rate $1/\sqrt{n}$ is nonetheless impossible, with a limit result similar to (4.3). Note that the condition in Corollary 4.4 involving $A \neq B\Lambda P$ is the minimum requirement for non-Gaussian error distributions to asymptotically distinguish F^A and F^B in the classical scenario (see Lemma 3.5).

Theorem 4.3 and Corollary 4.4 cover the case $\beta_n = o(1/\sqrt{n})$, in particular the case $\beta_n = n^{-\rho}$ for $\rho > 1/2$. We end this section with a result showing that, under some further regularity conditions, distinguishing $F^A_{\beta_n}$ and $F^B_{\beta_n}$ at rate $1/\sqrt{n}$ is possible when $0 < \rho < 1/2$.

Theorem 4.5. *Let* $\rho \in (0, 1/2)$ *and let* $\beta_n = n^{-\rho}$. *For all* $A \in \mathbb{M}(p,p)$, *define*

$$\Gamma_1(A) = \sum_{k=1}^{p} L_A\left(\zeta^{\otimes(k-1)} \otimes \frac{\xi - \zeta}{\|\xi - \zeta\|_\infty} \otimes \zeta^{\otimes(p-k)}\right). \tag{4.7}$$

If either $F^A \neq F^B$ *or* $F^A = F^B$ *and* $\Gamma_1(A) \neq \Gamma_1(B)$*, then*

$$\lim_{n\to\infty} P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) = 1. \tag{4.8}$$

As can be seen from the proof of Theorem 4.5, the measure $L_A(P_e(\beta)^{\otimes p})$ can be written as a polynomial of degree p in β, where the constant term corresponds to F^A and the first order term corresponds to $\Gamma_1(A)$, and similarly for $L_B(P_e(\beta)^{\otimes p})$. In this light, Theorem 4.5 shows that in the absence of a difference between the constant terms of $L_B(P_e(\beta)^{\otimes p})$ and $L_A(P_e(\beta)^{\otimes p})$, having different first order terms is a sufficient criterion for distinguishing $F^A_{\beta_n}$ and $F^B_{\beta_n}$ at rate $1/\sqrt{n}$.

5. Numerical experiments

In this section, we carry out numerical experiments related to the results in Section 4. To make our experiments feasible, we consider the scenario where $p = 2$, ζ is the standard normal distribution and ξ is the standard exponential distribution. We consider the two matrices

$$A = \begin{bmatrix} 1 & 0 \\ \alpha & \sqrt{1-\alpha^2} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \sqrt{1-\alpha^2} & \alpha \\ 0 & 1 \end{bmatrix},$$

where we set $\alpha = 0.4$. These two matrices are related to Example 1.1. Note that $AA^t = BB^t$ while $A \neq B\Lambda P$ for all diagonal Λ with $\Lambda^2 = I$ and all permutation matrices P. These properties makes A and B appropriate for evaluating the results of Section 4. Fix $\beta \in (0, 1)$ and let ε be a two-dimensional random variable with independent marginales and marginal distributions equal to $P_e(\beta) = \beta\xi + (1-\beta)\mathcal{N}$, where $\mathcal{N}$ denotes the standard normal distribution. The benefit of this setup is that the cumulative distribution functions F^A_β and F^B_β of $A\varepsilon$ and $B\varepsilon$ can be calculated in semi-analytical form, depending only on elementary functions and cumulative distribution functions for one-dimensional and two-dimensional normal distributions.

We consider numerical evaluation of the sequence in the left-hand side of (4.6) and approximation of its limit for $\beta_n = n^{-\rho}$ with varying $\rho > 0$. We use a Monte Carlo approximation to evaluate the probability. To be concrete, for fixed n and $\beta_n = n^{-\rho}$, we make the approximation

$$P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) \approx \frac{1}{N}\sum_{k=1}^{N} 1_{(X_k > c)} \tag{5.1}$$

for some fixed N, where (X_k) are independent and identically distributed variables with the same distribution as $\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty$. In order to simulate values from X_k, we first simulate n variables with distribution $A\epsilon$, this allows us to calculate the empirical cumulative distribution function in any point. Due to properties of empirical cumulative distribution functions, the supremum in

$\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty$ can be reduced to a finite one. However, for the purpose of reducing computational time, we are forced to approximate the finite maximum with a maximum over some fewer points. This means that our probability estimates in general will be biased downwards. Also because of time considerations, we restrict ourselves to using $N = 1000$ in (5.1).

Coupling the above with the semi-analytical form of the exact cumulative distribution function, we are capable of simulating values of X_k and evaluating the Monte Carlo estimate (5.1). Our numerical results are summarized in Figure 1.

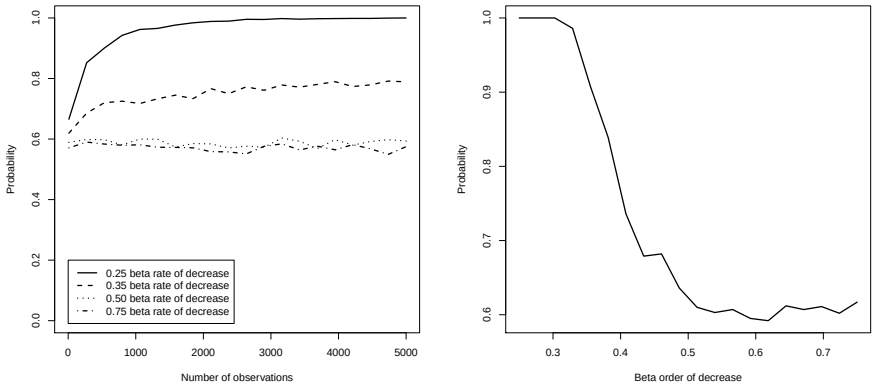

FIG 1. *Left: Estimates of* $P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c)$ *for* $\beta_n = n^{-\rho}$ *with* ρ *equal to* 0.25, 0.35, 0.50 *and* 0.75, *with* n *varying up to* 5000. *Right: Estimates of* $P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c)$ *for* $\beta_n = n^{-\rho}$ *with* ρ *varying from* 0.25 *to* 0.75. *Here,* $n = 50000$. *In both cases,* $c = 1$.

Theorem 4.3 states that for $\rho > 1/2$, the left hand side of (5.1) has a limit less than one. For ρ equal to 0.50 and 0.75, our numerical results in Figure 1 are in agreement with this.

On the other hand, for $\rho < 1/2$, Theorem 4.5 states that the limit of the left hand side of (5.1) equals 1. The results in Figure 1 indeed confirm this for $\rho = 0.25$. For ρ between 0.3 and 0.5, however, the results are less clear. We think this is caused by the fact that the sample size required to see the asymptotic behavior in Theorem 4.5 increases as ρ tends closer to 1/2. To see this, note that for large n, we have for some constant $K > 0$ that

$$P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) \approx P(\sqrt{n}\|F^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) \approx 1_{(\sqrt{n}K\beta_n > c)}, \quad (5.2)$$

since $||\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}||_\infty \approx ||F^A_{\beta_n} - F^B_{\beta_n}||_\infty$ and $||F^A_{\beta_n} - F^B_{\beta_n}||_\infty$ is approximately linear in β_n by Corollary 3.4. Therefore, $P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) \approx 1$ if it holds that $\sqrt{n}\beta_n > c/K$, corresponding to $n > \exp((\log c/K)/(1/2 - \rho))$. This indicates that the n required to detect the limiting value grows exponentially fast in $(1/2 - \rho)^{-1}$.

6. Discussion

We studied ICA models for error distributions which have independent and identically distributed coordinates following contaminated distributions. Our main theoretical contribution is Theorem 4.3, which shows that for $\beta_n = n^{-\rho}$ and $\rho \geq 1/2$, $\mathbb{F}^A_{\beta_n}$ is asymptotically as close to $F^B_{\beta_n}$ as to $F^A_{\beta_n}$ in the uniform norm whenever $F^A = F^B$. For contaminated Gaussian distributions, the requirement $F^A = F^B$ corresponds to $AA^t = BB^t$. Our results thus indicate that consistent estimation of A up to sign reversion and permutation of columns is not possible in this asymptotic scenario. In particular, causal inference as described in Example 1.1 (using LiNGAM) would suffer in such scenarios.

The proof of our main theoretical result, Theorem 4.3, rests on two partial results:

1. Lemma A.3, stating that when F_n is a sequence of cumulative distribution functions converging uniformly to F, and $\mathbb{F}_n$ is an empirical process based on n independent observations of variables with cumulative distribution function F_n, then $\sqrt{n}(\mathbb{F}_n - F_n)$ converges weakly in $\ell_\infty(\mathbb{R}^p)$.
2. Theorem 3.1, which is used to obtain that the convergence of the distribution functions F^A_β of is asymptotically linear in β as β tends to zero. This result, combined with Lemma 3.3, allows us to obtain an asymptotic bound on $\|F^A_\beta - F^B_\beta\|_\infty$ in Corollary 3.4.

In Theorem 4.5, we also considered the case of slower rates of decrease in the level of contamination, namely rates $n^{-\rho}$ for $0 < \rho < 1/2$. Our results here indicate that in such asymptotic scenarios, identifiability of the mixing matrix up to sign reversion and permutation of columns is possible, subject to some regularity conditions related to the Γ_1 signed measures of (4.7).

Our results are asymptotic in nature, considering limiting scenarios for a sequence of noise distributions. Theory on ICA applying such sequences is not common. One paper using similar methodologies is [22]. The authors of that paper are mainly interested in estimation of A^{-1} using nonparametric maximum likelihood. For this purpose, they introduce the log-concave ICA projection. In Theorem 5 of their paper, they essentially prove that the set of log-concave ICA projections for which the corresponding ICA model is identifiable is an open set, using sequences of ICA models (parametrized through sequences of probability measures). We hope that the present results as well as those of [22] indicate that considering the properties of ICA for various limiting scenarios of noise distributions may be an area for results on ICA not yet fully reaped.

Our results also open up new research questions, such as the following: Is it possible to characterize the matrices A and B such that the regularity condition $\Gamma_1(A) \neq \Gamma_2(B)$ of Theorem 4.5 holds? Also, together, Theorem 4.3 and Theorem 4.5 describe the behaviour of the empirical process $\mathbb{F}^A_{\beta_n}$ for asymptotic scenarios of the form $\beta_n = n^{-\rho}$ for $\rho > 0$, in particular describing the difficulty of using $\mathbb{F}^A_{\beta_n}$ to distinguish $F^A_{\beta_n}$ and $F^B_{\beta_n}$. Is it possible to obtain finite-sample bounds instead of limiting behaviours in these results? How do Theorem 4.3 and Theorem 4.5 translate into results on the ability of practical algorithms such as

the fastICA algorithm, see [11], to distinguish the correct mixing matrix? Is it possible to use similar techniques to analyze identifiability of the mixing matrix in asymptotic scenarios where the number of components p tends to infinity? Do the present results extend to cases where the coordinates of the error distributions are not contaminated normal distributions, or when the coordinates are not identically distributed?

Our results have been motivated by applications in causal inference. Besides linear SEMs with non-Gaussian noise as discussed in Example 1.1, there are other settings where the underlying causal structure is completely identifiable, such as non-linear SEMs with almost arbitrary additive noise and linear SEMs with additive Gaussian noise of equal variance, see e.g. [10] and [20], respectively. We may also ask whether one can use similar techniques to those presented here in order to study identifiability in these models when the structural equations are close to linear or the variance of the errors are close to equal, respectively. Also, we may consider the asymptotic behaviour of the model when we, instead of considering perturbations of the error distribution, consider perturbations of the additive nature of the noise, such that our SEM is defined by $X = f(X, \varepsilon)$, where $f(x) \approx Ax + \varepsilon$, corresponding to models with near-additive noise, and we let f tend to a function linear in ε?

In light of these open questions, our present results should be seen as a small step towards a better understanding of the identifiability of the mixing matrix for ICA for error distributions which are close to Gaussian but not Gaussian. We hope that this paper will lead to more work in this direction.

Appendix A: Proofs

A.1. Proofs for Section 3

Proof of Theorem 3.1. First note that we have $P_e(\beta) - \zeta = \beta(\xi - \zeta)$. Taking norms, this implies $\|P_e(\beta) - \zeta\|_\infty = \beta\|\xi - \zeta\|_\infty$ and

$$\frac{P_e(\beta) - \zeta}{\|P_e(\beta) - \zeta\|_\infty} = \frac{\xi - \zeta}{\|\xi - \zeta\|_\infty} = \nu. \tag{A.1}$$

We then also have $P_e(\beta) = \zeta + \beta\|\xi - \zeta\|_\infty \nu$.

We now analyze $P_e(\beta)^{\otimes p}$. For Borel subsets $C_1, \ldots, C_p$ of $\mathbb{R}$, we have

$$\begin{aligned} P_e(\beta)^{\otimes p}(C_1 \times \cdots \times C_p) &= (\zeta + \beta\|\xi - \zeta\|_\infty \nu)^{\otimes p}(C_1 \times \cdots \times C_p) \\ &= \prod_{k=1}^{p} (\zeta(C_k) + \beta\|\xi - \zeta\|_\infty \nu(C_k)) \\ &= \sum_{k=0}^{p} \beta^k \|\xi - \zeta\|_\infty^k \sum_{\alpha \in S_k} \prod_{i=1}^{p} \zeta(C_i)^{1-\alpha_i} \nu(C_i)^{\alpha_i}, \end{aligned} \tag{A.2}$$

where $S_k = \{\alpha \in \{0,1\}^p \,|\sum_{i=1}^p \alpha_i = k\}$, and the last equality follows since

$$\prod_{k=1}^{p}(a_k + \gamma b_k) = \sum_{k=0}^{p}\gamma^k \sum_{\alpha\in S_k}\prod_{i=1}^{p} a_i^{1-\alpha_i} b_i^{\alpha_i}, \qquad \text{for } a, b \in \mathbb{R}^p \text{ and } \gamma \in \mathbb{R}. \tag{A.3}$$

Defining $\mu_0 = \zeta$ and $\mu_1 = \nu$, we then obtain

$$P_e(\beta)^{\otimes p}(C_1 \times \cdots \times C_p) = \sum_{k=0}^{p}\beta^k \|\xi - \zeta\|_\infty^k \sum_{\alpha\in S_k}(\otimes_{i=1}^p \mu_{\alpha_i})(C_1 \times \cdots \times C_p). \tag{A.4}$$

Letting $\Gamma_k = \sum_{\alpha\in S_k} L_A(\otimes_{i=1}^p \mu_{\alpha_i})$, this yields

$$L_A(P_e(\beta)^{\otimes p}) = \sum_{k=0}^{p}\beta^k \|\xi - \zeta\|_\infty^k \Gamma_k. \tag{A.5}$$

Next, note that $\Gamma_0 = L_A(\zeta^{\otimes p})$, so that

$$\begin{aligned}\lim_{\beta\to 0}\frac{L_A(P_e(\beta)^{\otimes p}) - L_A(\zeta^{\otimes p})}{\|P_e(\beta) - \zeta\|_\infty} &= \lim_{\beta\to 0}\sum_{k=1}^{p}\beta^{k-1}\|\xi - \zeta\|_\infty^{k-1}\Gamma_k \\ &= \Gamma_1 = \sum_{k=1}^{p} L_A(\zeta^{\otimes(k-1)} \otimes \nu \otimes \zeta^{\otimes(p-k)}).\end{aligned} \tag{A.6}$$

In particular, this shows that for any $\eta > 0$,

$$\begin{aligned}\limsup_{\beta\to 0}\|F_\beta^A - F^A\|_\infty &= \limsup_{\beta\to 0}\|L_A(P_e(\beta)^{\otimes p}) - L_A(\zeta^{\otimes p})\|_\infty \\ &\le \limsup_{\beta\to 0}(1+\eta)\|\Gamma_1\|_\infty\|P_e(\beta) - \zeta\|_\infty \\ &\le \limsup_{\beta\to 0}(1+\eta)\|\Gamma_1\|_\infty\beta\|\xi - \zeta\|_\infty = 0,\end{aligned} \tag{A.7}$$

so F_β^A converges uniformly to F^A as β tends to zero. □

Proof of Lemma 3.2. Let $\beta \in (0,1)$ and let α be such that $\|Q(\beta) - \zeta\|_\infty = \alpha\beta$. Let $\xi = Q(1)$. We then have

$$\frac{Q(\beta) - \zeta}{\|Q(\beta) - \zeta\|_\infty} = \frac{Q(\beta) - \zeta}{\alpha\beta}, \tag{A.8}$$

while

$$\frac{Q(1) - \zeta}{\|Q(1) - \zeta\|_\infty} = \frac{\xi - \zeta}{\alpha}. \tag{A.9}$$

By our assumptions, the right-hand sides in (A.8) and (A.9) are equal. This implies $Q(\beta) = \beta\xi + (1-\beta)\zeta$. □

To prove Lemma 3.3, we first present a lemma relating the uniform norm of certain measures on $(\mathbb{R}^p, \mathcal{B}_p)$ to the uniform and total variation norms of some measures on $(\mathbb{R}, \mathcal{B})$.

Lemma A.1. *Let $\mu_1, \dots, \mu_p$ be signed measures on $(\mathbb{R}, \mathcal{B})$, and let $A \in \mathbb{M}(p,p)$. Then for any $i \in \{1, \dots, p\}$, it holds that*

$$\|L_A(\mu_1 \otimes \cdots \otimes \mu_p)\|_\infty \leq 2\|\mu_i\|_\infty \prod_{k \neq i}^{p} \|\mu_k\|_{tv}. \tag{A.10}$$

Proof. For any permutation $\pi : \{1, \dots, p\} \to \{1, \dots, p\}$ and corresponding permutation matrix P, we have $L_A(\mu_1 \otimes \cdots \otimes \mu_p) = L_{AP^{-1}}(\mu_{\pi(1)} \otimes \cdots \otimes \mu_{\pi(p)})$. Hence, it suffices to consider the case where $i = p$. Let $x \in \mathbb{R}^p$ and define $I_x = (-\infty, x_1] \times \cdots \times (-\infty, x_p]$. Then Fubini's theorem for signed measures yields

$$\begin{aligned}
&|L_A(\mu_1 \otimes \cdots \otimes \mu_p)(I_x)| \\
=&\left| \int \cdots \int 1_{I_x}(L_A(y)) \, \mathrm{d}\mu_p(y_p) \cdots \mathrm{d}\mu_1(y_1) \right| \\
\leq& \int \cdots \int \left| \int 1_{I_x}(L_A(y)) \, \mathrm{d}\mu_p(y_p) \right| \mathrm{d}|\mu_{p-1}|(y_{p-1}) \cdots \mathrm{d}|\mu_1|(y_1),
\end{aligned} \tag{A.11}$$

where we have also used the triangle inequality for integrals with respect to signed measures, which follows for example from Theorem 6.12 of [21]. We now analyze the innermost integral of (A.11). For fixed $y_1, \dots, y_{p-1}$, we have

$$\begin{aligned}
&\{y_p \in \mathbb{R} \mid 1_{I_x}(L_A(y)) = 1\} \\
=&\{y_p \in \mathbb{R} \mid \ \forall\, i \leq p : (Ay)_i \leq x_i\} \\
=& \cap_{i=1}^{p} \{y_p \in \mathbb{R} \mid a_{ip} y_p \leq x_i - (a_{i1} y_1 + \cdots + a_{i(p-1)} y_{p-1})\},
\end{aligned} \tag{A.12}$$

where a_{ij} is the (i,j)'th entry of A. Hence, $\{y_p \in \mathbb{R} \mid 1_{I_x}(L_A(y)) = 1\}$ is a finite intersection of intervals, and is therefore itself an interval. This yields

$$|\mu_p(\{y_p \in \mathbb{R} \mid 1_{I_x}(L_A(y)) = 1\})| \leq 2\|\mu_p\|_\infty. \tag{A.13}$$

This inequality is immediate when the interval is of the form $(-\infty, a]$ for some $a \in \mathbb{R}$. If the interval is of the form $[a, \infty)$, we have

$$\begin{aligned}
|\mu_p([a,\infty))| \leq& |\mu_p(\mathbb{R})| + |\mu_p(-\infty, a)| \\
=& \lim_{b \to \infty} |\mu_p((-\infty, b])| + |\mu_p((-\infty, a - 1/b])| \leq 2\|\mu_p\|_\infty,
\end{aligned} \tag{A.14}$$

and similarly for other types of intervals, whether bounded or unbounded, open, half-open or closed. Combining (A.11) and (A.13) yields

$$\begin{aligned}
&|L_A(\mu_1 \otimes \cdots \otimes \mu_p)(I_x)| \\
\leq& \int \cdots \int 2\|\mu_p\|_\infty \, \mathrm{d}|\mu_{p-1}|(y_{p-1}) \cdots \mathrm{d}|\mu_1|(y_1) = 2\|\mu_p\|_\infty \prod_{k=1}^{p-1} \|\mu_k\|_{tv}.
\end{aligned} \tag{A.15}$$

□

Proof of Lemma 3.3. By Lemma A.1, we have

$$\|L_A(\zeta^{\otimes(k-1)} \otimes \nu \otimes \zeta^{\otimes(p-k)})\|_\infty \leq 2\|\nu\|_\infty \|\zeta\|_{tv}^{p-1} = 2. \tag{A.16}$$

Applying the triangle inequality, we therefore obtain

$$\left\| \sum_{k=1}^{p} L_A \left(\zeta^{\otimes(k-1)} \otimes \nu \otimes \zeta^{\otimes(p-k)} \right) \right\|_\infty \leq 2p. \tag{A.17}$$

□

Proof of Lemma 3.5. **Proof of (1).** With ζ Gaussian with mean zero and variance σ^2, $L_A(\zeta^{\otimes p})$ is Gaussian with mean zero and variance $\sigma^2 AA^t$, and so the result is immediate for this case.

Proof of (3). Now consider the case where ζ is not a symmetric distribution. As $L_P(\zeta^{\otimes p}) = \zeta^{\otimes p}$ holds for any permutation matrix P, we obtain that if $A = BP$, then $L_A(\zeta^{\otimes p}) = L_B(\zeta^{\otimes p})$ and so $F^A = F^B$, proving one implication.

Conversely, assume that $F^A = F^B$, meaning that $L_A(\zeta^{\otimes p}) = L_B(\zeta^{\otimes p})$. As ζ is nondegenerate and non-Gaussian and A and B are invertible, Theorem 4 of [9] shows that $A = B\Lambda P$, where $\Lambda \in \mathbb{M}(p,p)$ is an invertible diagonal matrix and $P \in \mathbb{M}(p,p)$ is a permutation matrix. This yields

$$\zeta^{\otimes p} = L_{B^{-1}}(L_B(\zeta^{\otimes p})) = L_{B^{-1}}(L_A(\zeta^{\otimes p})) = L_{\Lambda P}(\zeta^{\otimes p}) = L_\Lambda(\zeta^{\otimes p}). \tag{A.18}$$

Now let Z be a random variable with distribution ζ. The above then yields that for all i, $\Lambda_{ii}Z$ and Z have the same distribution. In particular, $|\Lambda_{ii}||Z|$ and $|Z|$ have the same distribution, so $P(|Z| \leq z/|\Lambda_{ii}|) = P(|Z| \leq z)$ for all $z \in \mathbb{R}$. As Z is not almost surely zero, there is $z \neq 0$ such that $P(|Z| \leq z-\varepsilon) < P(|Z| \leq z+\varepsilon)$ for all $\varepsilon > 0$. This yields $|\Lambda_{ii}| = 1$. Next, let φ denote the characteristic function of Z. We then have $\varphi(\Lambda_{ii}\theta) = \varphi(\theta)$ for all $\theta \in \mathbb{R}$. As Z is not symmetric, there is a $\theta \in \mathbb{R}$ such that $\varphi(\theta) \neq \varphi(-\theta)$. Therefore, $\Lambda_{ii} = -1$ cannot hold, so we must have $\Lambda_{ii} = 1$. We conclude that Λ is the identity matrix and thus $A = BP$, as required.

Proof of (2). Finally, consider a symmetric probability measure ζ. It is then immediate that when Λ and P are as in the statement of the lemma, it holds that $L_{\Lambda P}(\zeta^{\otimes p}) = \zeta^{\otimes p}$ and thus $F^A = F^B$ whenever $A = B\Lambda P$. The converse implication follows as in the proof of (3). □

A.2. Proofs for Section 4

Before proving Theorem 4.2 and Theorem 4.3, we show a result on empirical processes. Recall that for a metric space (M,d), the ε-covering number $N(\varepsilon, M, d)$ is the minimum number of open balls of radius ε which is required to cover (M,d), see, e.g., Section 2.1.1 of [25].

Lemma A.2. *Fix a cumulative distribution function F. Define $\rho : \mathbb{R}^p \times \mathbb{R}^p$ by*

$$\rho(x,y) = \sqrt{F(x) + F(y) - 2F(x \wedge y)}, \tag{A.19}$$

and let $I_x = (-\infty, x_1] \times \cdots \times (-\infty, x_p]$. Let Z be a variable with cumulative distribution function F. Then, the following holds:

1. *ρ is a pseudometric.*
2. *$\rho(x,y) = \sqrt{E(1_{I_x}(Z) - 1_{I_y}(Z))^2}$.*
3. *$(\mathbb{R}^p, \rho)$ is totally bounded.*

Proof. First note that

$$\begin{aligned}\rho(x,y)^2 =& F(x) + F(y) - 2F(x \wedge y)\\ =& E1_{I_x}(Z) + E1_{I_y}(Z) - 2E1_{I_x}(Z)1_{I_y}(Z)\\ =& E(1_{I_x}(Z) - 1_{I_y}(Z))^2,\end{aligned} \tag{A.20}$$

proving claim (2). It is then immediate that ρ is a pseudometric, proving claim (1). Next, it holds that $(\mathbb{R}^p, \rho)$ is totally bounded if and only if $N(\varepsilon, \mathbb{R}^p, \rho)$ is finite for all positive ε. Let Q be the distribution corresponding to the cumulative distribution function F, and let $\mathcal{L}^2(\mathbb{R}^p, \mathcal{B}_p, Q)$ be the space of Borel measurable functions from $\mathbb{R}^p$ to $\mathbb{R}$ which are square-integrable with respect to Q. Let $\|\cdot\|_{2,Q}$ denote the usual seminorm on $\mathcal{L}^2(\mathbb{R}^p, \mathcal{B}_p, Q)$. Applying claim (2), it is immediate that

$$N(\varepsilon, \mathbb{R}^p, \rho) = N(\varepsilon, (1_{I_x})_{x\in\mathbb{R}^p}, \|\cdot\|_{2,Q}). \tag{A.21}$$

Combining Example 2.6.1 and Exercise 2.6.9 of [25], we find that $(1_{I_x})_{x\in\mathbb{R}^p}$ is a Vapnik-Cervonenkis (VC) subgraph class with VC dimension $p+1$. Furthermore, $(1_{I_x})_{x\in\mathbb{R}^p}$ has envelope function constant and equal to one. Therefore, Theorem 2.6.7 of [25] shows that $N(\varepsilon, (1_{I_x})_{x\in\mathbb{R}^p}, \|\cdot\|_{2,Q})$ and thus $N(\varepsilon, \mathbb{R}^p, \rho)$ is finite, and so $(\mathbb{R}^p, \rho)$ is totally bounded. □

Lemma A.3. *Let (F_n) be a sequence of cumulative distribution functions on $\mathbb{R}^p$, and let F be a cumulative distribution function on $\mathbb{R}^p$. Let $(X_{nk})_{1\le k\le n}$ be a triangular array such that for each n, $X_{n1}, \ldots, X_{nn}$ are independent with distribution F_n. Let $\mathbb{F}_n$ be the empirical distribution function of $X_{n1}, \ldots, X_{nn}$. If F_n converges uniformly to F, then $\sqrt{n}(\mathbb{F}_n - F_n)$ converges weakly in $\ell_\infty(\mathbb{R}^p)$ to an F-Gaussian field.*

Proof. For $x, y \in \mathbb{R}^p$ and $n \ge 1$, let $R_n(x,y) = F_n(x\wedge y) - F_n(x)F_n(y)$ and also define $R(x,y) = F(x\wedge y) - F(x)F(y)$. Let ρ be the pseudometric of Lemma A.2 corresponding to the cumulative distribution function F. Let Z_{nk} be the random field indexed by $\mathbb{R}^p$ given by $Z_{nk}(x) = 1_{I_x}(X_{nk})/\sqrt{n}$, where we as usual put $I_x = (-\infty, x_1] \times \cdots \times (-\infty, x_p]$. We then have

$$\begin{aligned}\sum_{k=1}^n Z_{nk}(x) - EZ_{nk}(x) =& \frac{1}{\sqrt{n}} \sum_{k=1}^n 1_{I_x}(X_{nk}) - F_n(x)\\ =& \sqrt{n}(\mathbb{F}_n(x) - F_n(x)).\end{aligned} \tag{A.22}$$

We will apply Theorem 2.11.1 of [25] to prove that $\sum_{k=1}^n Z_{nk} - EZ_{nk}$ and thus $\sqrt{n}(\mathbb{F}_n - F_n)$ converges weakly in $\ell_\infty(\mathbb{R}^p)$. We may assume without loss of generality that all variables are defined on a product probability space as described in Section 2.11.1 of [25], and as the fields (Z_{nk}) can be constructed using only

countably many variables, the measurability requirements in Theorem 2.11.1 of [25] can be ensured. In order to apply Theorem 2.11.1 of [25], first note that by Lemma A.2, $(\mathbb{R}^p, \rho)$ is totally bounded and so can be applied in Theorem 2.11.1 of [25]. Also, the covariance function of $\sum_{k=1}^n Z_{nk} - EZ_{nk}$ is

$$\begin{aligned}
&\operatorname{Cov}\left(\sum_{k=1}^n Z_{nk}(x) - EZ_{nk}(x), \sum_{k=1}^n Z_{nk}(y) - EZ_{nk}(y)\right)\\
=&\sum_{k=1}^n \sum_{i=1}^n EZ_{nk}(x)Z_{ni}(y) - EZ_{nk}(x)EZ_{ni}(y)\\
=&\frac{1}{n}\sum_{k=1}^n E1_{I_x}(X_{nk})1_{I_y}(X_{nk}) - E1_{I_x}(X_{nk})E1_{I_y}(X_{nk})\\
=&F_n(x \wedge y) - F_n(x)F_n(y) = R_n(x,y). \qquad \text{(A.23)}
\end{aligned}$$

Note that

$$\begin{aligned}
&|R(x,y) - R_n(x,y)|\\
\leq&|F(x \wedge y) - F_n(x \wedge y)| + |F(x)F(y) - F_n(x)F_n(y)|\\
\leq&|F(x \wedge y) - F_n(x \wedge y)| + |F(x) - F_n(x)| + |F_n(y) - F_n(y)|, \qquad \text{(A.24)}
\end{aligned}$$

so as F_n converges uniformly to F, R_n converges uniformly to R. Thus, the covariance functions of $\sum_{k=1}^n Z_{nk} - EZ_{nk}$ converge to R. Therefore, in order to apply Theorem 2.11.1 of [25], it only remains to confirm that the conditions of (2.11.2) in [25] hold. Fixing $\eta > 0$, we have

$$\begin{aligned}
\sum_{k=1}^n E\|Z_{nk}\|_\infty^2 1_{(\|Z_{nk}\|_\infty > \eta)} =&\frac{1}{n}\sum_{k=1}^n E1_{I_x}(X_{nk})1_{(1_{I_x}(X_{nk}) > \sqrt{n}\eta)}\\
\leq&P(1_{I_x}(X_{n1}) > \sqrt{n}\eta),
\end{aligned}$$

and so it is immediate that the first condition of (2.11.2) in [25] holds. Next, define $d_n^2(x,y) = \sum_{k=1}^n (Z_{nk}(x) - Z_{nk}(y))^2$. We then also have for $x, y \in \mathbb{R}^p$ that

$$d_n^2(x,y) = \frac{1}{n}\sum_{k=1}^n (1_{I_x}(X_{nk}) - 1_{I_y}(X_{nk}))^2, \qquad \text{(A.25)}$$

and therefore, $Ed_n(x,y)^2 = F_n(x) + F_n(y) - 2F_n(x \wedge y)$. Thus, $(x,y) \mapsto Ed_n(x,y)^2$ converges uniformly to ρ^2 on $\mathbb{R}^p \times \mathbb{R}^p$. Therefore, we conclude that for any sequence (δ_n) of positive numbers tending to zero, it holds for all $\eta > 0$ that

$$\begin{aligned}
\limsup_{n\to\infty} \sup_{x,y:\rho(x,y)\leq\delta_n} Ed_n^2(x,y) \leq&\limsup_{n\to\infty} \sup_{x,y:\rho(x,y)\leq\delta_n} \rho(x,y)^2\\
\leq&\limsup_{n\to\infty} \delta_n^2 = 0. \qquad \text{(A.26)}
\end{aligned}$$

Hence, the second condition of (2.11.2) in [25] holds. In order to verify the final condition of (2.11.2) in [25], first note that $d_n(x,y)^2 = E_{\mathbb{P}_n}(1_{I_x} - 1_{I_y})^2$ by

(A.25), where $E_{\mathbb{P}_n}$ denotes integration with respect to $\mathbb{P}_n$ and $\mathbb{P}_n$ is the empirical measure on $(\mathbb{R}^p, \mathcal{B}_p)$ in $X_{n1}, \ldots, X_{nn}$. Thus, $d_n(x, y)$ is the $\mathcal{L}^2(\mathbb{R}^p, \mathcal{B}_p, \mathbb{P}_n)$ distance between the mappings I_x and I_y, and so

$$N(\varepsilon, \mathbb{R}^p, d_n) = N(\varepsilon, (1_{I_x})_{x\in\mathbb{R}^p}, \|\cdot\|_{2,\mathbb{P}_n}) \le \sup_Q N(\varepsilon, (1_{I_x})_{x\in\mathbb{R}^p}, \|\cdot\|_{2,Q}), \quad \text{(A.27)}$$

where $\|\cdot\|_{2,Q}$ denotes the norm on $\mathcal{L}^2(\mathbb{R}^p, \mathcal{B}_p, Q)$ and the supremum is over all probability measures Q on $(\mathbb{R}^p, \mathcal{B}_p)$. Thus, the third condition of (2.11.2) in [25] is satisfied if only it holds that for all sequences (δ_n) of positive numbers tending to zero,

$$\lim_{n\to\infty} \int_0^{\delta_n} \sup_Q \sqrt{\log N(\varepsilon, (1_{I_x})_{x\in\mathbb{R}^p}, \|\cdot\|_{2,Q})}\, \mathrm{d}\varepsilon = 0. \quad \text{(A.28)}$$

However, Theorem 2.6.7 of [25] yields a constant $K > 0$ such that for $0 < \varepsilon < 1$,

$$\sup_Q N(\varepsilon, (1_{I_x})_{x\in\mathbb{R}^p}, \|\cdot\|_{2,Q}) \le K(p+1)(16e)^{p+1}\varepsilon^{-2p}. \quad \text{(A.29)}$$

As a consequence, again for $0 < \varepsilon < 1$,

$$\sup_Q \sqrt{\log N(\varepsilon, (1_{I_x})_{x\in\mathbb{R}^p}, \|\cdot\|_{2,Q})} \le \sqrt{\log K(p+1)(16e)^{p+1} - 2p\log\varepsilon}. \quad \text{(A.30)}$$

By elementary calculations, we obtain for $0 < c < d < 1$ and $a, b > 0$ that

$$\begin{aligned} &\int_c^d \sqrt{a - b\log x}\, \mathrm{d}x \\ &= \left[x\sqrt{a - b\log x} - \frac{e^{a/b}\sqrt{\pi b}}{2} \operatorname{erf}\left(\frac{\sqrt{a - b\log x}}{\sqrt{b}} \right) \right]_c^d, \end{aligned} \quad \text{(A.31)}$$

where erf denotes the error function, $\operatorname{erf}(x) = (2/\sqrt{\pi}) \int_0^x \exp(-y^2)\, \mathrm{d}y$. Therefore, we conclude that for all $0 < \eta < 1$, the mapping $x \mapsto \sqrt{a - b\log x}$ is integrable over $[0, \eta]$. Thus, (A.28) holds. Recalling (A.22), Theorem 2.11.1 of [25] now shows that $\sqrt{n}(\mathbb{F}_n - F_n)$ converges weakly in $\ell_\infty(\mathbb{R}^p)$. By uniqueness of the finite-dimensional distributions of the limit, we find that the limit is an F-Gaussian field. □

Proof of Theorem 4.2. By Lemma A.3 and the continuous mapping theorem, $\sqrt{n}\|\mathbb{F}_n^A - F^A\|_\infty$ converges weakly to $\|W\|_\infty$. Therefore, equation (4.3) follows. In order to prove equation (4.4), consider A and B such that $F^A \neq F^B$ and let $\|F^A - F^B\|_\infty = \alpha$. Whenever $\|\mathbb{F}_n^A - F^A\|_\infty \le \alpha/2$, the reverse triangle inequality yields

$$\begin{aligned} \|\mathbb{F}_n^A - F^B\|_\infty &= \|\mathbb{F}_n^A - F^A - (F^B - F^A)\|_\infty \\ &\ge \big|\|\mathbb{F}_n^A - F^A\|_\infty - \|F^B - F^A\|_\infty\big| \\ &= \big|\|\mathbb{F}_n^A - F^A\|_\infty - \alpha\big| \ge \alpha/2. \end{aligned} \quad \text{(A.32)}$$

Since $\lim_{n\to\infty} P(\|\mathbb{F}_n^A - F^A\|_\infty \leq \alpha/2) = 1$ by Lemma A.3, we obtain

$$\begin{aligned}
&\limsup_{n\to\infty} P(\sqrt{n}\|\mathbb{F}_n^A - F^B\|_\infty \leq c) \\
&\quad = \limsup_{n\to\infty} P(\|\mathbb{F}_n^A - F^B\|_\infty \leq c/\sqrt{n}, \|\mathbb{F}_n^A - F^A\|_\infty \leq \alpha/2) \\
&\quad \leq \limsup_{n\to\infty} P(\|\mathbb{F}_n^A - F^B\|_\infty \leq c/\sqrt{n}, \|\mathbb{F}_n^A - F^B\|_\infty \geq \alpha/2) = 0. \qquad \text{(A.33)}
\end{aligned}$$

Hence, $\lim_{n\to\infty} P(\sqrt{n}\|\mathbb{F}_n^A - F^B\|_\infty \leq c) = 0$ and so (4.4) holds. □

Proof of Theorem 4.3. By the triangle inequality, we have the inequalities

$$\begin{aligned}
&P(\sqrt{n}\|\mathbb{F}_{\beta_n}^A - F_{\beta_n}^A\|_\infty - \sqrt{n}\|F_{\beta_n}^B - F_{\beta_n}^A\|_\infty > c) \\
\leq &P(\sqrt{n}\|\mathbb{F}_{\beta_n}^A - F_{\beta_n}^B\|_\infty > c) \\
\leq &P(\sqrt{n}\|\mathbb{F}_{\beta_n}^A - F_{\beta_n}^A\|_\infty + \sqrt{n}\|F_{\beta_n}^B - F_{\beta_n}^A\|_\infty > c). \qquad \text{(A.34)}
\end{aligned}$$

Let $\eta > 0$. By Corollary 3.4, we can choose $N \geq 1$ such that for $n \geq N$,

$$\sqrt{n}\|F_{\beta_n}^B - F_{\beta_n}^A\|_\infty \leq 4p(1+\eta)\sqrt{n}\beta_n\|\xi - \zeta\|_\infty. \qquad \text{(A.35)}$$

By our assumptions, $\lim_n \sqrt{n}\beta_n = k$. Letting $\gamma > 0$, we then find for n large that

$$\sqrt{n}\|F_{\beta_n}^B - F_{\beta_n}^A\|_\infty \leq 4p(1+\eta)(k+\gamma)\|\xi - \zeta\|_\infty. \qquad \text{(A.36)}$$

For such n, the first inequality of (A.34) yields

$$\begin{aligned}
&P(\sqrt{n}\|\mathbb{F}_{\beta_n}^A - F_{\beta_n}^B\|_\infty > c) \\
\geq &P(\|\sqrt{n}(\mathbb{F}_{\beta_n}^A - F_{\beta_n}^A)\|_\infty > c + \sqrt{n}\|F_{\beta_n}^B - F_{\beta_n}^A\|_\infty) \\
\geq &P(\|\sqrt{n}(\mathbb{F}_{\beta_n}^A - F_{\beta_n}^A)\|_\infty > c + 4p(1+\eta)(k+\gamma)\|\xi - \zeta\|_\infty). \qquad \text{(A.37)}
\end{aligned}$$

Now recall from Theorem 3.1 that $F_{\beta_n}^A$ converges uniformly to F^A. Therefore, Lemma A.3 and the continuous mapping theorem show that $\sqrt{n}\|\mathbb{F}_{\beta_n}^A - F_{\beta_n}^A\|_\infty$ converges weakly to $\|W\|_\infty$. As a consequence, (A.37) yields

$$\begin{aligned}
&\liminf_{n\to\infty} P(\sqrt{n}\|\mathbb{F}_{\beta_n}^A - F_{\beta_n}^B\|_\infty > c) \\
\geq &P(\|W\|_\infty > c + 4p(1+\eta)(k+\gamma)\|\xi - \zeta\|_\infty). \qquad \text{(A.38)}
\end{aligned}$$

Letting η and then γ tend to zero, we obtain

$$\liminf_{n\to\infty} P(\sqrt{n}\|\mathbb{F}_{\beta_n}^A - F_{\beta_n}^B\|_\infty > c) \geq P(\|W\|_\infty > c + 4pk\|\xi - \zeta\|_\infty). \qquad \text{(A.39)}$$

Similarly, the second inequality of (A.34) yields

$$\begin{aligned}
&P(\sqrt{n}\|\mathbb{F}_{\beta_n}^A - F_{\beta_n}^B\|_\infty > c) \\
\leq &P(\|\sqrt{n}(\mathbb{F}_{\beta_n}^A - F_{\beta_n}^A)\|_\infty > c - \sqrt{n}\|F_{\beta_n}^B - F_{\beta_n}^A\|_\infty) \\
\leq &P(\|\sqrt{n}(\mathbb{F}_{\beta_n}^A - F_{\beta_n}^A)\|_\infty \geq c - 4p(1+\eta)(k+\gamma)\|\xi - \zeta\|_\infty), \qquad \text{(A.40)}
\end{aligned}$$

and by similar arguments as previously, we obtain

$$\limsup_{n\to\infty} P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) \leq P(\|W\|_\infty \geq c - 4pk\|\xi-\zeta\|_\infty). \tag{A.41}$$

Combining our results, we obtain (4.5). □

Proof of Corollary 4.4. As we have assumed that $P_e(\beta_n)$ is non-Gaussian, it follows from Lemma 3.5 that $F^A_\beta \neq F^B_\beta$, since $A \neq B\Lambda P$ for all diagonal Λ with $\Lambda^2 = I$ and all permutation matrices P. This shows (1). And as $AA^t = BB^t$ and ζ is Gaussian, Lemma 3.5 yields $F^A = F^B$, so Theorem 4.3 yields (2). □

Proof of Theorem 4.5. Note that for any $x \in \mathbb{R}^p$, we have

$$\begin{aligned} &P(\sqrt{n}\|\mathbb{F}^A_{\beta_n} - F^B_{\beta_n}\|_\infty > c) \\ \geq &P(\sqrt{n}|\mathbb{F}^A_{\beta_n}(x) - F^B_{\beta_n}(x)| > c) \\ = &P(|\sqrt{n}(\mathbb{F}^A_{\beta_n}(x) - F^A_{\beta_n}(x)) + \sqrt{n}(F^A_{\beta_n}(x) - F^B_{\beta_n}(x))| > c). \end{aligned} \tag{A.42}$$

We first consider the case $F^A \neq F^B$. Let $x \in \mathbb{R}^p$ be such that $F^A(x) \neq F^B(x)$. Then $\lim_n F^A_{\beta_n}(x) - F^B_{\beta_n}(x) \neq 0$, so $|\sqrt{n}(F^A_{\beta_n}(x) - F^B_{\beta_n}(x)|$ tends to infinity as n tends to infinity. By the central limit theorem, $\sqrt{n}(\mathbb{F}^A_{\beta_n}(x) - F^A_{\beta_n}(x))$ converges in distribution. Therefore, (A.42) yields the result.

Next, consider the case $F^A = F^B$ and $\Gamma_1(A) \neq \Gamma_1(B)$. Let $x \in \mathbb{R}^p$ be such that $\Gamma_1(A)(I_x) \neq \Gamma_1(B)(I_x)$. Similarly to the proof of Theorem 3.1, define $\mu_0 = \zeta$, $\mu_1 = (\xi-\zeta)/\|\xi-\zeta\|_\infty$, $S_k = \{\alpha \in \{0,1\}^p \,|\sum_{i=1}^p \alpha_i = k\}$ and also $\Gamma_k(A) = \sum_{\alpha\in S_k} L_A(\otimes_{i=1}^p \mu_{\alpha_i})$. Note that $\Gamma_k(A)$ with $k=1$ corresponds to (4.7). Then, we have

$$L_A(P_e(\beta)^{\otimes p}) = \sum_{k=0}^p \beta^k \|\xi-\zeta\|^k_\infty \Gamma_k(A), \tag{A.43}$$

see (A.5). In particular, we obtain

$$\begin{aligned} F^A_\beta(x) - F^B_\beta(x) =& L_A(P_e(\beta)^{\otimes p})(I_x) - L_B(P_e(\beta)^{\otimes p})(I_x) \\ =& \sum_{k=1}^p \beta^k \|\xi-\zeta\|^k_\infty (\Gamma_k(A)(I_x) - \Gamma_k(B)(I_x)), \end{aligned} \tag{A.44}$$

where we have used that $\Gamma_0(A) = \Gamma_0(B)$, since $F^A = F^B$. Since $\beta_n = n^{-\rho}$, we obtain

$$\sqrt{n}(F^A_{\beta_n}(x) - F^B_{\beta_n}(x)) = \sum_{k=1}^p n^{1/2-k\rho}\|\xi-\zeta\|^k_\infty (\Gamma_k(A)(I_x) - \Gamma_k(B)(I_x)). \tag{A.45}$$

As $\rho < 1/2$, $1/2-\rho > 0$. As $\|\xi-\zeta\|_\infty(\Gamma_1(A)(I_x) - \Gamma_1(B)(I_x)) \neq 0$, we conclude that as n tends to infinity, the term corresponding to $k=1$ in the above tends to infinity in absolute value. Since the right hand side of (A.45) is a sum with finitely many terms, where the remaining terms are of lower degree in n, we conclude that $|\sqrt{n}(F^A_{\beta_n}(x) - F^B_{\beta_n}(x))|$ tends to infinity as n tends to infinity. As in the previous case, since the ordinary central limit theorem shows that $\sqrt{n}(\mathbb{F}^A_{\beta_n}(x) - F^A_{\beta_n}(x))$ converges in distribution, (A.42) yields the result. □

Acknowledgements

We thank the anonymous reviewer and the editor for their comments and suggestions, which led to considerable improvements of the paper.

References

[1] S.-I. Amari and J.-F. Cardoso, *Blind source separation – semiparametric statistical approach*, IEEE Transactions on Signal Processing **45** (1997), no. 11, 2692–2700.

[2] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, *Face recognition by independent component analysis*, IEEE Transactions on Neural Networks **13** (2002), no. 6, 1450–1464.

[3] C. F. Beckmann and S. M. Smith, *Probabilistic independent component analysis for functional magnetic resonance imaging*, IEEE Transactions on Medical Imaging **23** (2004), no. 2, 137–152.

[4] A. Chen and P. J. Bickel, *Efficient independent component analysis*, Ann. Statist. **34** (2006), no. 6, 2825–2855.

[5] P. Comon, *Independent component analysis, a new concept?*, Signal Processing **36** (1994), 287–314.

[6] Pierre Comon and Christian Jutten, *Handbook of blind source separation: Independent component analysis and applications*, Elsevier, Oxford, 2010.

[7] M. Csörgő and L. Horváth, *A note on strong approximations of multivariate empirical processes*, Stochastic Process. Appl. **28** (1988), no. 1, 101–109.

[8] R. M. Dudley, *Weak convergences of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces*, Illinois J. Math. **10** (1966), 109–126.

[9] J. Eriksson and V. Koivunen, *Identifiability, separability and uniqueness of linear ICA models*, IEEE Signal Processing Letters **11** (2004), no. 7, 601–604.

[10] P. O. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf, *Nonlinear causal discovery with additive noise models*, Advances in Neural Information Processing Systems 21 (NIPS), MIT Press, 2009, pp. 689–696.

[11] A Hyvärinen, *Fast and robust fixed-point algorithms for independent component analysis*, IEEE Transactions on Neural Networks **10** (1999), no. 3, 626–634.

[12] ———, *Independent component analysis: Recent advances*, Phil. Trans. Roy. Soc. Ser. A **371** (2013), 1–19.

[13] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, Wiley-Blackwell, New York, 2001.

[14] P. Ilmonen and D. Paindaveine, *Semiparametrically efficient inference based on signed ranks in symmetric independent component models*, Ann. Statist. **39** (2011), no. 5, 2448–2476.

[15] T. P. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T.-W. Lee, and T. J. Sejnowski, *Imaging brain dynamics using independent component analysis*, Proceedings of the IEEE **89** (2001), no. 7, 1107–1122.

[16] D. Khoshnevisan, *Multiparameter processes*, Springer Monographs in Mathematics, Springer-Verlag, New York, 2002.
[17] M. A. Lifshits, *Gaussian random functions*, Mathematics and its Applications, vol. 322, Kluwer Academic Publishers, Dordrecht, 1995.
[18] M. Novey and T. Adah, *Complex ICA by negentropy maximization*, IEEE Transactions on Neural Networks **19** (2008), no. 4, 596–609.
[19] E. Ollila, H.-J. Kim, and V. Koivunen, *Compact Cramér-Rao bound expression for independent component analysis*, IEEE Transactions on Signal Processing **56** (2008), no. 4, 1421–1428.
[20] J. Peters and P. Bühlmann, *Identifiability of Gaussian structural equation models with same error variances*, Biometrika, to appear (2012), 1–11.
[21] W. Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987.
[22] R. J. Samworth and M. Yuan, *Independent component analysis via nonparametric maximum likelihood estimation*, Ann. Statist. **40** (2012), 2973–3002.
[23] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen, *A linear non-Gaussian acyclic model for causal discovery*, J. Mach. Learn. Res. **7** (2006), 2003–2030.
[24] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen, *DirectLiNGAM: a direct method for learning a linear non-Gaussian structural equation model*, J. Mach. Learn. Res. **12** (2011), 1225–1248.
[25] A. W. van der Vaart and J. A. Wellner, *Weak convergence and empirical processes*, Springer Series in Statistics, Springer-Verlag, New York, 1996.
[26] R. Vigario, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, *Independent component approach to the analysis of EEG and MEG recordings*, IEEE Transactions on Biomedical Engineering **47** (2000), no. 5, 589–593.
[27] M. Yuan, *On the identifiability of additive index models*, Stat. Sinica **21** (2011), 1901–1911.