

Minimum Distance Index for BSS, Generalization, Interpretation and Asymptotics

Niko Lietzén
Aalto University
School of Science

Joni Virta
Aalto University
School of Science &
University of Turku

Klaus Nordhausen
Institute of Statistics &
Mathematical Methods
in Economics,
Vienna University
of Technology

Pauliina Ilmonen
Aalto University
School of Science

Abstract

We consider complex valued linear blind source separation, where the signal dimension might be smaller than the dimension of the observable data vector. In order to measure the success of the signal separation, we propose an extension of the minimum distance index and establish its properties. Interpretations for the index are derived through connections to signal-to-noise ratios and correlations. The interpretations are novel also for the real valued original case. In addition, we consider the asymptotic behavior of the extended minimum distance index. This paper is an invited extended version of the paper presented at the CDAM 2019 conference.

Keywords: blind source separation, performance indices, asymptotic properties.

1. Introduction

In several application areas, including, e.g., signal processing, economics and biomedical applications, it is assumed that one observes vectors that are linear mixtures of latent source variables. See, e.g., [Comon and Jutten \(2010\)](#) for an introduction. The corresponding model is referred to as linear blind source separation (BSS) model. Usually, the goal in BSS is to recover the latent unobservable variables. That is, the objective is to estimate the corresponding mixing (or unmixing) matrix.

Several algorithms for solving the linear BSS problem have been presented in the literature, see for example [Ilmonen and Paindaveine \(2011\)](#); [Matteson and Tsay \(2017\)](#); [Miettinen, Ilner, Nordhausen, Oja, Taskinen, and Theis \(2016\)](#); [Nordhausen \(2014\)](#); [Risk, Matteson, Ruppert, Eloyan, and Caffo \(2014\)](#); [Virta, Li, Nordhausen, and Oja \(2017\)](#). As theoretical results, including asymptotic distributions, are often not presented, different estimators are compared in simulation studies. To enable comparisons, performance indices are needed. Popular indices considered in the literature include the Amari index, interference to signal ratio (ISR), mean square error (MSE) and the minimum distance index (MDI), see [Nordhausen, Ollila, and Oja \(2011\)](#) for a comparative study. MDI has the advantage of being affine invariant. Moreover, its

asymptotic behavior is known in the case of real valued square mixing matrices, see [Ilmonen, Nordhausen, Oja, and Ollila \(2010, 2012\)](#). In this paper, our focus is on the MDI for complex valued non-square matrices.

This paper is an invited extension of the paper presented at the CDAM 2019 conference [Lietzén, Virta, Nordhausen, and Ilmonen \(2019\)](#). In the conference paper, we extended the MDI to the case where only some of the sources, the signals, are of interest to us, and the index measures how well these signals are recovered. Furthermore, in order to cover applications such as biomedical image processing, where the observed signals can be complex valued, we worked under the assumption of complex valued signals. A real valued version of this extension was proposed already in [Virta, Nordhausen, and Oja \(2016\)](#). However, no theoretical justification for its properties was given in [Virta *et al.* \(2016\)](#). Similarly, a complex valued version of the regular MDI (where no distinction between the signal and the noise was made) was introduced in [Lietzén, Nordhausen, and Ilmonen \(2016\)](#). In this paper, we review the results presented in the aforementioned conference paper [Lietzén *et al.* \(2019\)](#). In particular, Lemma 2, Theorem 3 and Theorem 8, and the corresponding proofs, were already given in [Lietzén *et al.* \(2019\)](#). In this paper, we extend the theoretical results of [Lietzén *et al.* \(2019\)](#) by considering the limiting distribution of the non-square complex valued MDI. Moreover, to highlight the applicability and interpretation of the MDI, we present an interesting new image data example.

2. Linear complex valued BSS model

Let \mathbf{x} be an observable \mathbb{C}^p -valued random vector. Assume that,

$$\mathbf{x} = \mathbf{\Omega}\mathbf{z} = \mathbf{\Omega} \begin{pmatrix} \mathbf{z}_1^\top & \mathbf{z}_0^\top \end{pmatrix}^\top, \quad (1)$$

where $\mathbf{\Omega}$ is a full-rank $\mathbb{C}^{p \times p}$ -matrix and the latent \mathbb{C}^p -vector $\mathbf{z} = \begin{pmatrix} \mathbf{z}_1^\top & \mathbf{z}_0^\top \end{pmatrix}^\top$, consists of two parts. In this linear complex valued BSS model, the \mathbb{C}^d -vector \mathbf{z}_1 contains the signals of interest. Furthermore, the \mathbb{C}^{d_0} -vector \mathbf{z}_0 , $d + d_0 = p$, contains uninteresting noise.

Given Model (1), the objective is to find a $\mathbb{C}^{d \times p}$ transformation matrix $\mathbf{\Gamma}$, such that the transformed vector corresponds to the signal variables, $\mathbf{\Gamma}\mathbf{x} = \mathbf{z}_1$, up to some class of transformations, by using only the information contained in the observable \mathbf{x} . Note that, usually the transformation matrix $\mathbf{\Gamma}$ is not unique. For example, in independent component analysis (ICA), where the latent signals are assumed to be stochastically independent, the transformation matrix is unique up to heterogeneous scaling, permutations and phase-shifts of the rows. Thus in general, we have no guarantees that two separate IC estimation procedures estimate the same population quantities, which is something we need to consider when measuring performances. The minimum distance index discussed in the next section solves the issue by measuring how close the corresponding gain matrix $\mathbf{G} = \mathbf{\Gamma}\mathbf{\Omega}$ is to the matrix $\mathcal{J}_{d,p} = \begin{pmatrix} \mathbf{I}_d & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{d \times p}$, up to heterogeneous scaling, permutations and phase-shifts of the rows.

3. Complex valued non-square MDI

In this section, we present the definition of the complex valued non-square MDI. We start by considering the concept of equivalent matrices.

Definition 1. Let \sim be a relation on $\mathbb{C}^{d \times p}$, defined by $\mathbf{A} \sim \mathbf{B} \iff \mathbf{A} = (\mathbf{P}\mathbf{J}\mathbf{D})\mathbf{B}$, for some $\mathbf{D} \in \mathcal{D}^d$, $\mathbf{J} \in \mathcal{J}^d$ and $\mathbf{P} \in \mathcal{P}^d$, where \mathcal{P}^d is the set of $\mathbb{R}^{d \times d}$ permutation matrices, \mathcal{D}^d is the set of $\mathbb{R}^{d \times d}$ diagonal matrices with strictly positive real valued diagonal entries and \mathcal{J}^d is the set of $\mathbb{C}^{d \times d}$ diagonal matrices with diagonal entries of the form $\exp(i\theta_1), \dots, \exp(i\theta_d)$, where i is the imaginary unit. We use \mathcal{C}^d to denote the set defined as

$$\mathcal{C}^d = \{\mathbf{C} \in \mathbb{C}^{d \times d} \mid \mathbf{C} = \mathbf{P}\mathbf{J}\mathbf{D} : \mathbf{P} \in \mathcal{P}^d, \mathbf{J} \in \mathcal{J}^d, \mathbf{D} \in \mathcal{D}^d\}.$$

Lemma 2. *The relation \sim is an equivalence relation on $\mathbb{C}^{d \times p}$.*

We use the equivalence relation \sim to partition $\mathbb{C}^{d \times p}$ into equivalence classes and denote the set of matrices that are equivalent to the matrix \mathbf{A} as $\mathcal{C}_{\mathbf{A}} = \{\mathbf{B} \in \mathbb{C}^{d \times p} \mid \mathbf{A} \sim \mathbf{B}\}$.

We next consider the shortest squared distance between the equivalence class $\mathcal{C}_{\mathbf{A}}$ and the matrix $\mathfrak{J}_{d,p} = (\mathbf{I}_d \ \mathbf{0}) \in \mathbb{R}^{d \times p}$. The corresponding optimization problem is,

$$\text{minimize} \quad \|\mathbf{P}\mathbf{J}\mathbf{D}\mathbf{A} - \mathfrak{J}_{d,p}\|_{\mathbb{F}}^2 \quad \text{s.t.} \quad \mathbf{P} \in \mathcal{P}^d, \quad \mathbf{J} \in \mathcal{J}^d \quad \text{and} \quad \mathbf{D} \in \mathcal{D}^d, \quad (2)$$

where $\|\cdot\|_{\mathbb{F}}$ is the Frobenius norm. Note that the diagonal elements of matrices belonging to \mathcal{D}^d are in the open interval $(0, \infty)$, and thus, the optimization problem does not necessarily have a minimizer. This leads us to define the shortest distance between the equivalence class $\mathcal{C}_{\mathbf{A}}$ and the matrix $\mathfrak{J}_{d,p}$ using infimum as follows,

$$\varrho(\mathbf{A}, \mathfrak{J}_{d,p}) = \inf_{\mathbf{B} \in \mathcal{C}_{\mathbf{A}}} \|\mathbf{B} - \mathfrak{J}_{d,p}\|_{\mathbb{F}} = \inf_{\mathbf{C} \in \mathcal{C}^d} \|\mathbf{C}\mathbf{A} - \mathfrak{J}_{d,p}\|_{\mathbb{F}}. \quad (3)$$

The squared distance $[\varrho(\mathbf{A}, \mathfrak{J}_{d,p})]^2$ can be converted into a more applicable form, see the following theorem.

Theorem 3. *Let $\mathbf{A} \in \mathbb{C}^{d \times p}$ and $\mathfrak{J}_{d,p} = (\mathbf{I}_d \ \mathbf{0}) \in \mathbb{R}^{d \times p}$, $d \leq p$. Furthermore, let $\tilde{\mathbf{A}}_{jk} = |\mathbf{A}_{jk}|^2 / \sum_{h=1}^p |\mathbf{A}_{jh}|^2$, if \mathbf{A} has at least one nonzero element in row j and $\tilde{\mathbf{A}}_{jk} = 0$ if \mathbf{A} has only zeros in row j . Then, the squared distance $[\varrho(\mathbf{A}, \mathfrak{J}_{d,p})]^2$ defined in Eq. (3), coincides with,*

$$d - \max_{\mathbf{P} \in \mathcal{P}^d} \left\{ \text{Trace} \left(\mathbf{P}\tilde{\mathbf{A}} \right) \right\},$$

where the trace of a non-square matrix is the sum of its main-diagonal elements.

Proof of Theorem 3. We find the greatest lower bound by allowing the matrix \mathbf{D} to have zeros on the diagonal. We combine the optimization variables \mathbf{J} and \mathbf{D} by optimizing over a variable $\mathbf{L} \in \mathcal{L}^d$, where \mathcal{L}^d is the set of all $\mathbb{C}^{d \times d}$ diagonal matrices. Since the Frobenius norm is orthogonally invariant, the corresponding objective function f can be reformulated as follows,

$$f(\mathbf{P}, \mathbf{L}, \mathfrak{J}_{d,p}) = \|\mathbf{P}\mathbf{L}\mathbf{A} - \mathfrak{J}_{d,p}\|_{\mathbb{F}}^2 = \left\| \mathbf{P} \left(\mathbf{L}\mathbf{A} - \mathbf{P}^{\top} \mathfrak{J}_{d,p} \right) \right\|_{\mathbb{F}}^2 = \left\| \mathbf{L}\mathbf{A} - \mathbf{P}^{\top} \mathfrak{J}_{d,p} \right\|_{\mathbb{F}}^2.$$

Next, write $\mathbf{A} = \mathbf{V} + i\mathbf{W}$ and $\mathbf{L} = \mathbf{Q} + i\mathbf{R}$, where $\mathbf{Q}, \mathbf{R}, \mathbf{V}, \mathbf{W}$ have real valued entries. Now, since \mathbf{L} is a diagonal matrix, we obtain the following form for $f(\mathbf{P}, \mathbf{L}, \mathfrak{J}_{d,p})$:

$$\sum_{j=1}^d \sum_{k=1}^p \left[\mathbf{Q}_{jj}^2 (\mathbf{V}_{jk}^2 + \mathbf{W}_{jk}^2) + \mathbf{R}_{jj}^2 (\mathbf{V}_{jk}^2 + \mathbf{W}_{jk}^2) \right] - 2 \sum_{j,k=1}^d (\mathbf{Q}_{jj} \mathbf{V}_{jk} - \mathbf{R}_{jj} \mathbf{W}_{jk}) \mathbf{P}_{kj} + d.$$

In the formula above, we have applied the constraints that the off-diagonal elements of \mathbf{Q} and \mathbf{R} are zero. Thus, the only remaining constraint is that \mathbf{P} is a permutation matrix. We proceed to verify the Karush-Kuhn-Tucker (KKT) necessary conditions.

Assume that \mathbf{A} has ℓ rows that have at least one nonzero element. The $d - \ell$ rows that contain only zeros give no contribution to the objective function and we can without loss of generality permute \mathbf{A} such that the ℓ first rows of \mathbf{A} are the ones with at least one nonzero element. The partial derivatives with respect to the first ℓ diagonal elements of \mathbf{Q} and \mathbf{R} are,

$$\begin{aligned} \frac{\partial f(\mathbf{P}, \mathbf{L}, \mathfrak{J}_{d,p})}{\partial \mathbf{Q}_{jj}} &= \sum_{k=1}^p 2\mathbf{Q}_{jj} (\mathbf{V}_{jk}^2 + \mathbf{W}_{jk}^2) - 2 \sum_{k=1}^d \mathbf{V}_{jk} \mathbf{P}_{kj}, \\ \frac{\partial f(\mathbf{P}, \mathbf{L}, \mathfrak{J}_{d,p})}{\partial \mathbf{R}_{jj}} &= \sum_{k=1}^p 2\mathbf{R}_{jj} (\mathbf{V}_{jk}^2 + \mathbf{W}_{jk}^2) + 2 \sum_{k=1}^d \mathbf{W}_{jk} \mathbf{P}_{kj}, \end{aligned}$$

from which we can derive the solution candidates \mathbf{Q}'_{jj} and \mathbf{R}'_{jj} , which are optimal by the convexity of the corresponding optimization problem,

$$\mathbf{Q}'_{jj} = \frac{\sum_{k=1}^d \mathbf{V}_{jk} \mathbf{P}_{kj}}{\sum_{k=1}^p (\mathbf{V}_{jk}^2 + \mathbf{W}_{jk}^2)} \quad \text{and} \quad \mathbf{R}'_{jj} = \frac{-\sum_{k=1}^d \mathbf{W}_{jk} \mathbf{P}_{kj}}{\sum_{k=1}^p (\mathbf{V}_{jk}^2 + \mathbf{W}_{jk}^2)}, \quad j \leq \ell.$$

Since the last $d-\ell$ rows of \mathbf{A} do not contribute to the objective function, we set $\mathbf{Q}'_{jj} = \mathbf{L}'_{jj} = 0$ and $\mathbf{P}_{jj} = 1$, when $j > \ell$. Using the property $\mathbf{V}_{jk}^2 + \mathbf{W}_{jk}^2 = |\mathbf{A}_{jk}|^2$ and since a permutation matrix has exactly one nonzero element 1 in every row and column, we can reformulate the objective function for $\mathbf{L}' = \mathbf{Q}' + i\mathbf{R}'$ as follows,

$$f(\mathbf{P}, \mathbf{L}', \mathcal{J}_{d,p}) = d - \sum_{j=1}^{\ell} \frac{|\mathbf{A}_{j,\pi(\mathbf{P},j)}|^2 \mathbf{P}_{\pi(\mathbf{P},j),j}}{\sum_{k=1}^p |\mathbf{A}_{j,k}|^2} - \sum_{j=\ell+1}^d 0 = d - \sum_{j=1}^d \tilde{\mathbf{A}}_{j,\pi(\mathbf{P},j)}$$

such that $\pi : \mathcal{P}^d \times \{1, \dots, d\} \rightarrow \{1, \dots, d\} : \mathbf{P} \times j \mapsto c_j$, where c_j is the row-index for the only nonzero element of the permutation matrix \mathbf{P} in the column j . An equivalent optimization problem is then to find a permutation for the rows of $\tilde{\mathbf{A}}$, such that the sum of the main-diagonal elements is maximized, that is,

$$\min_{\mathbf{P} \in \mathcal{P}^d} \left\{ d - \sum_{j=1}^d \tilde{\mathbf{A}}_{j,\pi(\mathbf{P},j)} \right\} \iff d - \max_{\mathbf{P} \in \mathcal{P}^d} \left\{ \text{Trace}(\mathbf{P}\tilde{\mathbf{A}}) \right\}.$$

□

As in [Ilmonen *et al.* \(2012\)](#), the trace maximization problem in [Theorem 3](#) can be seen as a linear sum assignment problem (LSAP). It can be solved, e.g., by the Hungarian method [Papadimitriou and Steiglitz \(1982\)](#). In practice, the LSAP can be solved using, e.g., the `solve_LSAP` function given in the statistical software R, [R Core Team \(2018\)](#), package `clue`, [Hornik \(2005\)](#).

We next present the minimum distance index (MDI) for non-square complex valued matrices. Note that this is an extension to the cases presented in [Ilmonen *et al.* \(2010, 2012\)](#); [Lietzén *et al.* \(2016\)](#).

Definition 4. Let $\mathbf{\Omega}$ be the mixing matrix in [Model \(1\)](#), let $\hat{\mathbf{\Gamma}}$ be a corresponding unmixing estimate and let $\hat{\mathbf{G}} = \hat{\mathbf{\Gamma}}\mathbf{\Omega}$ be $\mathbb{C}^{d \times p}$ -valued. The minimum distance index (MDI) for the estimate $\hat{\mathbf{\Gamma}}$ is given by,

$$\text{MD}(\hat{\mathbf{\Gamma}}) = \frac{\varrho(\hat{\mathbf{G}}, \mathcal{J}_{d,p})}{\sqrt{d}} = \frac{1}{\sqrt{d}} \inf_{\mathbf{C} \in \mathbb{C}^d} \left\| \mathbf{C}\hat{\mathbf{G}} - (\mathbf{I}_d \ \mathbf{0}) \right\|_F.$$

Remark 5. Note that, the non-square MDI is applicable beyond [Model \(1\)](#). It can, for example, be applied for assessing the performance when the mixing matrix $\mathbf{\Omega} \in \mathbb{C}^{p \times q}$, $q \leq p$.

Note that, in order to apply the MDI, the mixing matrix $\mathbf{\Omega}$ needs to be known. We next show that MDI satisfies properties that enable interpretations and fair comparisons.

Theorem 6. Let $\mathbf{A} \in \mathbb{C}^{d \times p}$. Then, $\text{MD}(\mathbf{A}) \in [0, 1]$ and the following properties are satisfied,

(i) $\text{MD}(\mathbf{A}) = 0 \iff \mathbf{A}\mathbf{\Omega} \sim (\mathbf{I}_d \ \mathbf{0}) = \mathcal{J}_{d,p}$,

(ii) $\text{MD}(\mathbf{A}) = 1 \iff \exists \mathbf{B} \in \mathbb{C}^{d \times (p-d)} : \mathbf{A}\mathbf{\Omega} = (\mathbf{0} \ \mathbf{B})$,

(iii) the function $f : [0, 1] \rightarrow \mathbb{R}$, $f(c) = \varrho((\mathbf{I}_d \ \mathbf{0}) + c \cdot \text{off}(\mathbf{A}), \mathcal{J}_{d,p})$, is increasing in $c \in [0, 1]$ for all matrices \mathbf{A} that satisfy $|\mathbf{A}_{jk}| < 1$ when $j \neq k$. (The function $\text{off}(\cdot)$ here sets the main-diagonal elements of its argument equal to zero.)

Proof of Theorem 6. Let $\mathbf{G} = \mathbf{A}\mathbf{\Omega}$ and let $\tilde{\mathbf{G}}_{jk} = |\mathbf{G}_{jk}|^2 / \sum_{h=1}^p |\mathbf{G}_{jh}|^2$, if \mathbf{G} has at least one nonzero element in row j and $\tilde{\mathbf{G}}_{jk} = 0$ if \mathbf{G} has only zeros in row j . By Theorem 3, we have that $\text{MD}^2(\mathbf{A}) = 1 - \frac{1}{d} \max_{\mathbf{P}} \{\text{Trace}(\mathbf{P}\tilde{\mathbf{G}})\}$, where the elements of $\tilde{\mathbf{G}}$ are between zero and one. Hereby, the minimal value for $\text{MD}^2(\mathbf{A})$ is obtained, when $\mathbf{P}\tilde{\mathbf{G}}$ is an identity matrix, and consequently, $\text{Trace}(\mathbf{P}\tilde{\mathbf{G}}) = d$. In addition, the maximal value for $\text{MD}^2(\mathbf{A})$ is obtained, when the first d columns of $\tilde{\mathbf{G}}$ contain only zeros, and consequently, $\text{Trace}(\mathbf{P}\tilde{\mathbf{G}}) = 0$. Thus, it follows that $\text{MD}(\mathbf{A}) \in [0, 1]$. Next, we proceed to verify properties (i)–(iii).

First, assume that $\mathbf{G} \sim \mathcal{J}_{d,p}$. This gives us that $\tilde{\mathbf{G}}$ is equal, up to a permutation, to $\mathcal{J}_{d,p}$. The maximal trace is then achieved by the permutation that places the only nonzero elements 1 of the matrix $\tilde{\mathbf{G}}$ to the main-diagonal. Thus, $\max_{\mathbf{P}} \{\text{Trace}(\mathbf{P}\tilde{\mathbf{G}})\} = d$, which implies $\text{MD}(\mathbf{A}) = 0$. Next, assume that $\text{MD}(\mathbf{A}) = 0$. This assumption gives us that $\max_{\mathbf{P}} \{\text{Trace}(\mathbf{P}\tilde{\mathbf{G}})\} = d$. The corresponding trace can be d only when there exists an element $\tilde{\mathbf{G}}_{jk}$ in every row j such that $k \leq d$ and $\tilde{\mathbf{G}}_{jk} = 1$. Hereby, there has to be exactly one nonzero element in each of the first d columns of \mathbf{G} and the nonzero elements have to be on different rows. Consequently, \mathbf{G} is equivalent to $\mathcal{J}_{d,p}$. Thus, property (i) holds.

For property (ii), first assume that $\mathbf{G} = (\mathbf{0} \ \mathbf{B})$. Then, the main-diagonal of $\tilde{\mathbf{G}}_{jk}$ contains only zeros, regardless of the permutation and regardless of the matrix \mathbf{B} . Thus, $\max_{\mathbf{P}} \{\text{Trace}(\mathbf{P}\tilde{\mathbf{G}})\} = 0$, which gives us that $\text{MD}(\mathbf{A}) = 1$. For the only if part, assume that $\text{MD}(\mathbf{A}) = 1$. This assumption gives us that $\text{Trace}(\mathbf{P}\tilde{\mathbf{G}}) = 0$ for the optimal \mathbf{P} and consequently for every other \mathbf{P} as well. Thus, $\mathbf{G} = (\mathbf{0} \ \mathbf{B})$, where $\mathbf{B} \in \mathbb{C}^{d \times (p-d)}$ is arbitrary. Hereby, property (ii) holds.

For the final property (iii), assume that $0 \leq c_1 \leq c_2 \leq 1$. The requirement that the absolute values of the off-diagonal elements of \mathbf{A} are less than one ensures that the permutation matrix that maximizes the trace is the identity matrix. Then,

$$[f(c_2)]^2 - [f(c_1)]^2 = \sum_{k=1}^d \left(\frac{-1}{c_2 \sum_{j \neq k}^p |\mathbf{A}_{kj}|^2 + 1} + \frac{1}{c_1 \sum_{j \neq k}^p |\mathbf{A}_{kj}|^2 + 1} \right) \geq 0,$$

that is, the function f is increasing under the given conditions. □

Remark 7. Note that Theorem 6 and its proof are very similar to (Lietzén *et al.* 2019, Theorem 6). However, (Lietzén *et al.* 2019, Theorem 6) contains an error that was corrected in Theorem 6 and its proof.

4. Two interpretations for MDI

Theorem 6 reveals that the endpoints of the MDI-range $[0, 1]$ correspond to completely successful and unsuccessful signal separations, respectively. However, it is not entirely clear how much the separation result improves when the index decreases within the interval, say, from 0.2 to 0.1. To better interpret the MDI-values in practice, we next express MDI through two better-known measures of estimation accuracy, signal-to-noise ratio (SNR) and correlation.

Assume Model (1) and let, without loss of generality, the identifiability constraint $\text{Cov}(\mathbf{z}) = \mathbf{I}_p$ hold. Let further $\hat{\mathbf{\Gamma}} \in \mathbb{C}^{d \times p}$ be an estimated unmixing matrix for the d signals of interest, $\mathbf{z}_1 = (z_1, \dots, z_d)$. The estimates of the signals are then $\hat{\mathbf{z}}_1 = (\hat{z}_1, \dots, \hat{z}_d)^\top = \hat{\mathbf{G}}\mathbf{z}$, where $\hat{\mathbf{G}} = \hat{\mathbf{\Gamma}}\mathbf{\Omega} \in \mathbb{C}^{d \times p}$ is the gain matrix. We define the SNR and the absolute correlation of the j th signal to be,

$$\text{SNR}_j = \frac{\text{Var}(\hat{z}_j)}{\text{Var}(\hat{z}_j - z_j)} \quad \text{and} \quad |\text{Cor}_j| = |\text{Cor}(\hat{z}_j, z_j)|.$$

The above form of SNR is as defined in Gonzales and Woods (2001). Both SNR_j and $|\text{Cor}_j|$ measure the accuracy with which we estimate the j th signal. They have the respective ranges $[0, \infty)$ and $[0, 1]$, with the upper endpoints corresponding to a perfect estimation of the j th signal.

Our next result establishes a connection between MDI, SNR_j and $|\text{Cor}_j|$ under two constant contamination models for the gain matrix $\hat{\mathbf{G}}$: a homogeneous model where the contamination is equal for each of the elements, and a heterogeneous model where the contaminations are allowed to differ element-by-element but (for mathematical tractability) we require the diagonal elements of $\hat{\mathbf{G}}$ to be uncontaminated. These contamination models mimic situations, where there are estimation errors of different severities in the estimation of the unmixing matrix. Denote in the following the arithmetic and harmonic means of the values $c_1, \dots, c_d > 0$ by $\mathbf{A}(c_j) = \mathbf{A}(c_1, \dots, c_d)$ and $\mathbf{H}(c_j) = \mathbf{H}(c_1, \dots, c_d)$, respectively.

Theorem 8. *Assume Model (1) with $\text{Cov}(\mathbf{z}) = \mathbf{I}_p$.*

- i) *Under the homogeneous contamination, $\hat{\mathbf{G}} = \mathfrak{I}_{d,p} + \varepsilon \mathbf{1}_{d,p}$, where $\mathbf{1}_{d,p}$ is a $d \times p$ matrix full of ones and $\varepsilon \in \mathbb{C}, \text{Re}(\varepsilon) > -1/2$, we have*

$$\text{MD}^2(\hat{\mathbf{G}}) = \left(\frac{p-1}{p} \right) \text{SNR}_j^{-1} = 1 - |\text{Cor}_j|^2.$$

- ii) *Under the heterogenous contamination, $\hat{\mathbf{G}} = \mathfrak{I}_{d,p} + \mathbf{B}$, where $\mathbf{B} \in \mathbb{C}^{d \times p}$ has zero main diagonal and $|\mathbf{B}_{jk}| < 1, j \neq k$, we have*

$$\text{MD}^2(\hat{\mathbf{G}}) = \mathbf{H}(\text{SNR}_1, \dots, \text{SNR}_d)^{-1} = 1 - \mathbf{A}(|\text{Cor}_1|^2, \dots, |\text{Cor}_d|^2).$$

Proof of Theorem 8. Under the homogeneous contamination, we have that $\text{MD}^2(\hat{\mathbf{G}}) = 1 - \frac{1}{d} \max_{\mathbf{P}} \{\text{Trace}(\mathbf{P}\hat{\mathbf{G}})\}$, where each diagonal element of $\hat{\mathbf{G}}$ equals $|1 + \varepsilon|^2 / [|1 + \varepsilon|^2 + (p-1)|\varepsilon|^2]$ and each of its off-diagonal elements equals $|\varepsilon|^2 / [|1 + \varepsilon|^2 + (p-1)|\varepsilon|^2]$. The assumption $\text{Re}(\varepsilon) > -1/2$ guarantees that \mathbf{I}_d is the unique solution to the maximization problem over $d \times d$ permutation matrices, yielding $\text{MD}^2(\hat{\mathbf{G}}) = [(p-1)|\varepsilon|^2] / [|1 + \varepsilon|^2 + (p-1)|\varepsilon|^2]$. Now, by $\text{Cov}(\mathbf{z}) = \mathbf{I}_p$, the signal-to-noise ratio SNR_j of the j th signal $\hat{z}_j = \hat{\mathbf{G}}_{j1}z_1 + \dots + \hat{\mathbf{G}}_{jd}z_p$, is

$$\text{SNR}_j = \frac{\text{Var}(\hat{z}_j)}{\text{Var}(\hat{z}_j - z_j)} = \frac{\sum_{k=1}^p |\hat{\mathbf{G}}_{jk}|^2}{|\hat{\mathbf{G}}_{jj} - 1|^2 + \sum_{k \neq j}^p |\hat{\mathbf{G}}_{jk}|^2} = \left(\frac{p-1}{p} \right) \text{MD}^{-2}(\hat{\mathbf{G}}). \quad (4)$$

Similarly, the second-order statistics $\text{Var}(z_j) = 1$, $\text{Var}(\hat{z}_j) = |1 + \varepsilon|^2 + (p-1)|\varepsilon|^2$ and $\text{Cov}(\hat{z}_j, z_j) = 1 + \varepsilon$ imply that the squared absolute correlation for the j th signal has the value

$$|\text{Cor}_j|^2 = \frac{|\text{Cov}(\hat{z}_j, z_j)|^2}{\text{Var}(\hat{z}_j)\text{Var}(z_j)} = \frac{|1 + \varepsilon|^2}{|1 + \varepsilon|^2 + (p-1)|\varepsilon|^2} = 1 - \text{MD}^2(\hat{\mathbf{G}}),$$

concluding the proof for the homogeneous contamination.

The claim is shown for the heterogeneous contamination in exactly analogous manner, and we list only the values of some of the key quantities. The squared MDI is now $\text{MD}^2(\hat{\mathbf{G}}) = 1 - (1/d) \sum_{j=1}^d (1 + \sum_{k \neq j}^p |\hat{\mathbf{G}}_{jk}|^2)^{-1}$, and, using the form (2) for the signal-to-noise ratio, we obtain $1 - (\text{SNR}_j)^{-1} = (1 + \sum_{k \neq j}^p |\hat{\mathbf{G}}_{jk}|^2)^{-1}$, which together prove the result for the SNR. For the correlation, we have the values $\text{Var}(z_j) = 1$, $\text{Var}(\hat{z}_j) = 1 + \sum_{k \neq j}^p |\hat{\mathbf{G}}_{jk}|^2$ and $\text{Cov}(\hat{z}_j, z_j) = 1$, which together imply that $|\text{Cor}_j|^2 = (1 + \sum_{k \neq j}^p |\hat{\mathbf{G}}_{jk}|^2)^{-1}$, yielding the final part of the claim. \square

In Table 1, we demonstrate the results of Theorem 8 under the heterogeneous contamination (where the conversion formula does not depend on the dimensionality p), with the SNR-values expressed in the standard decibel scale through the transformation $\text{SNR} \mapsto 10 \log_{10}(\text{SNR})$.

Based on the table, the value $\text{MDI} = 0.1$ is already an indicator of a very satisfactory separation, corresponding to an average SNR of 20 dB and to an average absolute correlation of 0.995. Next, we visually demonstrate the MDI in the context of image separation.

Table 1: The average SNRs and absolute correlations for particular MDI-values.

MDI	0.1	0.01	0.001	0.0001
$\mathbb{H}(\text{SNR}_j)$	20 dB	40 dB	60 dB	80 dB
$\sqrt{\mathbb{A}(\text{Cor}_j ^2)}$	$1 - 0.5 \cdot 10^{-2}$	$1 - 0.5 \cdot 10^{-4}$	$1 - 0.5 \cdot 10^{-6}$	$1 - 0.5 \cdot 10^{-8}$.

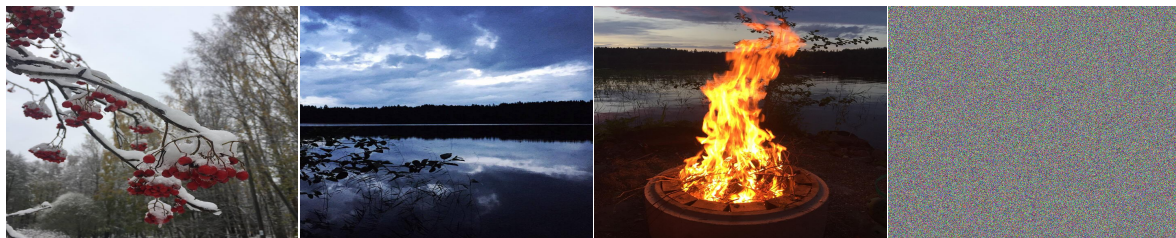
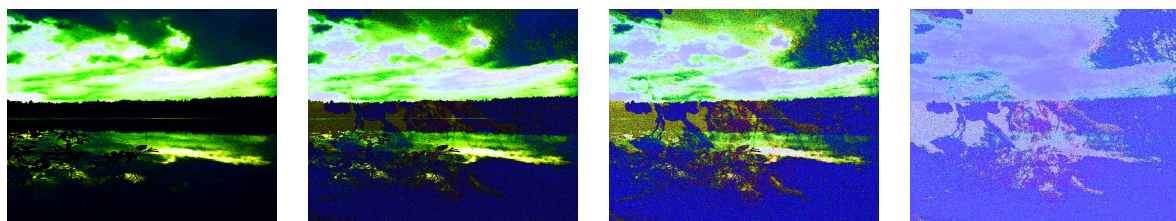


Figure 1: Original images.



(a) MD = 0.01

(b) MD = 0.25

(c) MD = 0.5

(d) MD = 0.85

Figure 2: Illustration of homogeneous contamination of the second image in Figure 1 with MDI values. The images are linear combinations of the images presented in Figure 1.

We consider four unmixed color images, presented in Figure 1. The example was conducted using the statistical software R, [R Core Team \(2018\)](#). We have chosen images with colors on the surface of the RGB cube, that is, at least one of the three values that define the color of a single pixel has either the maximal or minimal value. We can then find a transformation between the RGB cube surface and the complex plane. The transformation is not defined on a single point and is bijective everywhere else. The images in this example are chosen such that none of their pixels are on the non-bijective point. We have then homogeneously contaminated the images and calculated the corresponding MD index values for them. As an example, we present the contaminated second image, with the corresponding MDI values, in Figure 2. The contamination is conducted as in Theorem 8 part i), which ensures that the optimal permutation is the identity matrix, and thus the order of the images remains unchanged. As the contamination is homogeneous, the omitted images illustrate similar results. Note that the actual color in Figure 2 does not play a role in calculating the MDI as the color corresponds to the phase-shift and MDI is invariant with respect to scale, complex phase and order of the components.

5. Asymptotic behavior

In this section we provide the limiting distribution of the minimum distance index under general settings. In addition, we provide an example of the asymptotic properties, when we have \sqrt{n} -consistency and asymptotic normality.

For an n -indexed sequence $\hat{\mathbf{B}}_{(n)}$ of $\mathbb{C}^{d \times p}$ -matrices, we use $\hat{\mathbf{B}}_{(n)} = \mathcal{O}_p(1)$ to denote that the univariate sequences that correspond to the elements of $\hat{\mathbf{B}}_{(n)}$ are uniformly tight, i.e., bounded in probability. Additionally, we use $\hat{\mathbf{B}}_{(n)} = o_p(1)$ to denote that a sequence converges in

probability to a zero matrix.

Under general settings, finding the asymptotic distribution of the MDI can be complicated. In the following theorem, we present a direct asymptotic link between a sequence $\hat{\mathbf{A}}_{(n)}$ and the squared distance $[\varrho(\hat{\mathbf{A}}_{(n)}, \mathfrak{I}_{d,p})]^2$. By applying the theorem, one can usually significantly simplify the problem of finding the asymptotic distribution of the corresponding MDI.

Theorem 9. *Let $\hat{\mathbf{A}}_{(n)}$ be an n -indexed sequence of $\mathbb{C}^{d \times p}$ -matrices that satisfies $\gamma_n(\hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p}) = \mathcal{O}_p(1)$, where $(\gamma_n)_{n \in \mathbb{N}}$ is a real-valued increasing sequence that tends to infinity as $n \rightarrow \infty$. Then, as $n \rightarrow \infty$,*

$$\gamma_n^2 \left[\varrho \left(\hat{\mathbf{A}}_{(n)}, \mathfrak{I}_{d,p} \right) \right]^2 = \gamma_n^2 \left\| \text{off} \left(\hat{\mathbf{A}}_{(n)} \right) \right\|_{\text{F}}^2 + o_p(1),$$

where $\varrho(\hat{\mathbf{A}}_{(n)}, \mathfrak{I}_{d,p})$ is defined in Eq. (3) and the function $\text{off}(\cdot)$ sets the main-diagonal elements of its argument equal to zero.

Proof of Theorem 9. Note that, the assumption $\gamma_n(\hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p}) = \mathcal{O}_p(1)$ implies that $\hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} = o_p(1)$. Let,

$$\hat{\mathbf{C}}_{(n)} = \arg \min_{\mathbf{C} \in \overline{\mathcal{C}^d}} \left\| \mathbf{C} \hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right\|_{\text{F}},$$

where $\overline{\mathcal{C}^d} = \{ \mathbf{C} \in \mathbb{C}^{d \times d} \mid \mathbf{C} = \mathbf{P} \mathbf{J} \mathbf{D} : \mathbf{P} \in \mathcal{P}^d, \mathbf{J} \in \mathcal{J}^d, \mathbf{D} \in \overline{\mathcal{D}^d} \}$, and $\overline{\mathcal{D}^d}$ is the set of $\mathbb{R}^{d \times d}$ diagonal matrices with non-negative diagonal entries. We can then write,

$$\varrho \left(\hat{\mathbf{A}}_{(n)}, \mathfrak{I}_{d,p} \right) = \inf_{\mathbf{C} \in \overline{\mathcal{C}^d}} \left\| \mathbf{C} \hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right\|_{\text{F}} = \left\| \hat{\mathbf{C}}_{(n)} \hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right\|_{\text{F}}.$$

Similarly as in the proof of Theorem 3, we can split $\hat{\mathbf{C}}_{(n)}$ into a permutation matrix part $\hat{\mathbf{P}}_{(n)} \in \mathcal{P}^d$ and a diagonal matrix part $\hat{\mathbf{L}}_{(n)} \in \mathcal{L}^d$, such that $\hat{\mathbf{C}}_{(n)} = \hat{\mathbf{P}}_{(n)} \hat{\mathbf{L}}_{(n)}$. Since $\hat{\mathbf{A}}_{(n)}$ converges in probability to $\mathfrak{I}_{d,p}$, we obtain that $\hat{\mathbf{C}}_{(n)} - \mathbf{I}_d = o_p(1)$, see (Arcones 1998, Theorem 1). This further implies that $\hat{\mathbf{P}}_{(n)} - \mathbf{I}_d = o_p(1)$ and $\hat{\mathbf{L}}_{(n)} - \mathbf{I}_d = o_p(1)$. In addition, since the number of possible permutation matrices is finite, we have that $\gamma_n(\hat{\mathbf{P}}_{(n)} - \mathbf{I}_d) = o_p(1)$. For additional details, see the beginning of the proof of (Ilmonen *et al.* 2012, Theorem 4.2), and note that similar arguments can be applied here.

We can then apply the following factorization,

$$\begin{aligned} \gamma_n \left(\hat{\mathbf{C}}_{(n)} \hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right) &= \gamma_n \left[\left(\hat{\mathbf{P}}_{(n)} - \mathbf{I}_d \right) \left(\hat{\mathbf{L}}_{(n)} - \mathbf{I}_d \right) \left(\hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right) \right. \\ &\quad + \left(\hat{\mathbf{P}}_{(n)} - \mathbf{I}_d \right) \left(\hat{\mathbf{L}}_{(n)} - \mathbf{I}_d \right) \mathfrak{I}_{d,p} + \left(\hat{\mathbf{P}}_{(n)} - \mathbf{I}_d \right) \left(\hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right) \\ &\quad + \left(\hat{\mathbf{P}}_{(n)} - \mathbf{I}_d \right) \mathfrak{I}_{d,p} + \left(\hat{\mathbf{L}}_{(n)} - \mathbf{I}_d \right) \left(\hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right) \\ &\quad \left. + \left(\hat{\mathbf{L}}_{(n)} - \mathbf{I}_d \right) \mathfrak{I}_{d,p} + \left(\hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right) \right] \\ &= \gamma_n \left(\hat{\mathbf{L}}_{(n)} - \mathbf{I}_d \right) \mathfrak{I}_{d,p} + \gamma_n \left(\hat{\mathbf{A}}_{(n)} - \mathfrak{I}_{d,p} \right) + o_p(1). \end{aligned} \tag{5}$$

In the following, we denote, $\hat{a}_{jk} := [\hat{\mathbf{A}}_{(n)}]_{j,k}$, $\hat{S}_j := \sum_{k=1}^p |\hat{a}_{jk}|^2$ and $\hat{p}_{jk} := [\hat{\mathbf{P}}_{(n)}]_{j,k}$. Furthermore, we denote the complex conjugate of a as a^* . Note that $\hat{S}_j = 1 + o_p(1)$, and that $\gamma_n |\hat{a}_{jk}|^2 = \gamma_n (\hat{a}_{jk} \hat{a}_{jk}^*) = o_p(1)$, when $j \neq k$. Recall the form of the entries of the matrix $\hat{\mathbf{L}}_{(n)}$ from Theorem 3. The main-diagonal element (j, j) of Eq. (5) can then be expressed as

follows,

$$\begin{aligned}
& \gamma_n \left[\frac{\sum_{k=1}^d \hat{a}_{jk}^* \hat{p}_{kj}}{\hat{S}_j} - \frac{1}{\hat{S}_j} + \frac{1}{\hat{S}_j} - 1 + (\hat{a}_{jj} - 1) \right] + o_p(1) \\
&= \frac{\gamma_n}{\hat{S}_j} \left[(\hat{a}_{jj}^* \hat{p}_{jj} - 1) + \left(1 - \sum_{k=1}^p |\hat{a}_{jk}|^2 \right) + \hat{S}_j (\hat{a}_{jj} - 1) \right] + o_p(1) \\
&= \frac{\gamma_n}{\hat{S}_j} \left[(\hat{a}_{jj}^* - 1) (\hat{p}_{jj} - 1) + (\hat{p}_{jj} - 1) + (\hat{a}_{jj}^* - 1) + (1 - \hat{a}_{jj} \hat{a}_{jj}^*) + \hat{S}_j (\hat{a}_{jj} - 1) \right] + o_p(1) \\
&= \frac{\gamma_n}{\hat{S}_j} \left[(\hat{a}_{jj}^* - 1) + (1 - \hat{a}_{jj} \hat{a}_{jj}^*) + \hat{S}_j (\hat{a}_{jj} - 1) \right] + o_p(1) \\
&= \frac{\gamma_n}{\hat{S}_j} \left[(\hat{a}_{jj}^* - 1) (1 - \hat{a}_{jj}) + (1 - \hat{a}_{jj}) + \hat{S}_j (\hat{a}_{jj} - 1) \right] + o_p(1) \\
&= \frac{1}{\hat{S}_j} (1 - \hat{S}_j) \gamma_n (1 - \hat{a}_{jj}) + o_p(1) = \mathcal{O}_p(1) o_p(1) \mathcal{O}_p(1) + o_p(1) = o_p(1).
\end{aligned}$$

The above representation uses the formulation for $\hat{\mathbf{L}}_{(n)}$ that assumes at least one nonzero entry in every row of $\hat{\mathbf{A}}_{(n)}$. As we have that $\hat{\mathbf{A}}_{(n)}$ converges in probability to $\mathfrak{J}_{d,p}$, and since the theorem is an asymptotic result, it suffices to only consider the case of every row containing at least one nonzero element.

Thus, by using the continuous mapping theorem and the properties of the Frobenius norm,

$$\begin{aligned}
\gamma_n^2 \left[\varrho \left(\hat{\mathbf{A}}_{(n)}, \mathfrak{J}_{d,p} \right) \right]^2 &= \left\| \gamma_n \left(\hat{\mathbf{C}}_{(n)} \hat{\mathbf{A}}_{(n)} - \mathfrak{J}_{d,p} \right) \right\|_{\text{F}}^2 = \left\| \gamma_n \cdot \text{off} \left(\hat{\mathbf{A}}_{(n)} \right) + o_p(1) \right\|_{\text{F}}^2 \\
&= \gamma_n^2 \left\| \text{off} \left(\hat{\mathbf{A}}_{(n)} \right) \right\|_{\text{F}}^2 + o_p(1).
\end{aligned}$$

□

Hereby, the asymptotic distribution of the squared MDI is obtained by scaling both sides of Theorem 9 with $1/d$. Remarkably, Theorem 9 gives that the asymptotic distribution of MDI depends only on the off-diagonal elements of the sequence $\hat{\mathbf{A}}_{(n)}$.

We next present the widely considered case when $\gamma_n = \sqrt{n}$ and when $\hat{\mathbf{A}}_{(n)}$ is asymptotically normal.

Corollary 10. *Let $\hat{\mathbf{A}}_{(n)} = \hat{\mathbf{V}}_{(n)} + i \hat{\mathbf{W}}_{(n)}$ be an n -indexed sequence of $\mathbb{C}^{d \times p}$ -matrices that satisfy,*

$$\sqrt{n} \begin{pmatrix} \text{vec} \left(\hat{\mathbf{V}}_n - \mathfrak{J}_{d,p} \right) \\ \text{vec} \left(\hat{\mathbf{W}}_n \right) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathbf{Z},$$

where $\xrightarrow[n \rightarrow \infty]{\mathcal{D}}$ denotes convergence in distribution and $\mathbf{Z} \sim \mathcal{N}_{2dp}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Z}})$. Then,

$$n \left(\frac{\varrho \left(\hat{\mathbf{A}}_n, \mathfrak{J}_{d,p} \right)}{\sqrt{d}} \right)^2 \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y,$$

where Y is distributed as $d^{-1} \sum_{j=1}^{\ell} \lambda_j x_j$, where x_1, \dots, x_k are independent and $x_j \sim \chi^2(1)$, $\forall j \in \{1, \dots, \ell\}$, and $\lambda_1, \dots, \lambda_{\ell}$ are the ℓ nonzero eigenvalues (including all algebraic multiplicities) of $\boldsymbol{\Sigma}$, defined as,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{I}_{pd} - \mathbf{D}_{d,p} & 0 \\ 0 & \mathbf{I}_{pd} - \mathbf{D}_{d,p} \end{pmatrix} \boldsymbol{\Sigma}_{\mathbf{Z}} \begin{pmatrix} \mathbf{I}_{pd} - \mathbf{D}_{d,p} & 0 \\ 0 & \mathbf{I}_{pd} - \mathbf{D}_{d,p} \end{pmatrix},$$

where $\mathbf{D}_{d,p} = \text{diag}(\text{vec}(\mathfrak{J}_{d,p}))$.

Proof of Corollary 10. Since,

$$n \left\| \text{off} \left(\hat{\mathbf{A}}_{(n)} \right) \right\|_{\text{F}}^2 = n \sum_{j=1}^d \sum_{\substack{k=1 \\ j \neq k}}^p \left| \left[\hat{\mathbf{A}}_{(n)} \right]_{j,k} \right|^2 = n \sum_{j=1}^d \sum_{\substack{k=1 \\ j \neq k}}^p \left[\left(\left[\hat{\mathbf{V}}_{(n)} \right]_{j,k} \right)^2 + \left(\left[\hat{\mathbf{W}}_{(n)} \right]_{j,k} \right)^2 \right],$$

we can apply Theorem 9 and (Tan 1977, Theorem 3.1), which together with the idempotency of $\mathbf{I}_{pd} - \mathbf{D}_{d,p}$ give that $n(\varrho(\hat{\mathbf{A}}_n, \mathfrak{J}_{d,p})/\sqrt{d})^2$ converges in distribution, as $n \rightarrow \infty$, to a random variable distributed as $d^{-1} \sum_{j=1}^{\ell} \lambda_j x_j$, where x_1, \dots, x_{ℓ} are independent and $x_j \sim \chi^2(1)$, $\forall j \in \{1, \dots, \ell\}$, and $\lambda_1, \dots, \lambda_{\ell}$ are the ℓ nonzero eigenvalues of $\mathbf{\Sigma}$. \square

Examples of estimators that satisfy the assumptions of the sequence $\hat{\mathbf{A}}_{(n)}$ in Theorem 9, and in particular, Corollary 10, are given in, e.g., Imonen and Paindaveine (2011); Imonen (2013); Miettinen *et al.* (2016).

6. Final remarks

In this paper, we considered asymptotic behavior of the MDI for complex valued non-square matrices. The MDI is a powerful tool in controlled performance comparisons, when the mixing matrix $\mathbf{\Omega}$ is known. As stated in Remark 5, the extended MDI presented in this paper can be straightforwardly implemented to the undercomplete BSS case when $\mathbf{\Omega}$ has more rows than columns. Note that in the undercomplete case, the mixing matrix can always be converted to a square matrix by adding zeros.

In the future, we will explore whether MDI could be extended to nonlinear BSS. Additionally, an extension to the case of added noise, $\mathbf{x} = \mathbf{\Omega}\mathbf{z} + \boldsymbol{\varepsilon}$, will be considered in future work.

Acknowledgements: The work of N. Lietzén was supported by the Emil Aaltonen Foundation Grant 190135 N. The research of J. Virta was supported by the Academy of Finland Grant 321883. The work of K. Nordhausen was supported by the Austrian Science Fund (FWF) Grant number P31881-N32. The authors would like to thank Katariina Kilpinen for providing the photographs to the example in Section 4. We would like to thank the anonymous referee for insightful comments, which helped to improve the paper.

References

- Arcones MA (1998). “Asymptotic Theory for M-Estimators over a Convex Kernel.” *Economic Theory*, **14**(4), 387–422.
- Comon P, Jutten C (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, Oxford.
- Gonzales RC, Woods RE (2001). *Digital Image Processing Second Edition*. Prentice Hall, Upper Saddle River, New Jersey.
- Hornik K (2005). “A CLUE for CLUster Ensembles.” *Journal of Statistical Software*, **14**(12).
- Imonen P (2013). “On Asymptotic Properties of the Scatter Matrix Based Estimates for Complex Valued Independent Component Analysis.” *Statistics & Probability Letters*, **83**(4), 1219–1226.
- Imonen P, Nordhausen K, Oja H, Ollila E (2010). “A New Performance Index for ICA: Properties, Computation and Asymptotic Analysis.” In *International Conference on Latent Variable Analysis and Signal Separation*, pp. 229–236. Springer.

- Ilmonen P, Nordhausen K, Oja H, Ollila E (2012). “On Asymptotics of ICA Estimators and Their Performance Indices.” *arXiv Preprint arXiv:1212.3953*.
- Ilmonen P, Paindaveine D (2011). “Semiparametrically Efficient Inference Based on Signed Ranks in Symmetric Independent Component Models.” *Annals of Statistics*, **39**(5), 2448–2476.
- Lietzén N, Nordhausen K, Ilmonen P (2016). “Minimum Distance Index for Complex Valued ICA.” *Statistics & Probability Letters*, **118**, 100–106.
- Lietzén N, Virta J, Nordhausen K, Ilmonen P (2019). “Minimum Distance Index for Non-Square Complex Valued Mixing Matrices.” In *Proceedings of the XII International Conference on Computer Data Analysis & Modeling 2019 (CDAM)*, pp. 79–86. Belarusian State University Publishing House.
- Matteson DS, Tsay RS (2017). “Independent Component Analysis via Distance Covariance.” *Journal of the American Statistical Association*, **112**(518), 623–637.
- Miettinen J, Illner K, Nordhausen K, Oja H, Taskinen S, Theis FJ (2016). “Separation of Uncorrelated Stationary Time Series Using Autocovariance Matrices.” *Journal of Time Series Analysis*, **37**(3), 337–354.
- Nordhausen K (2014). “On Robustifying Some Second Order Blind Source Separation Methods for Nonstationary Time Series.” *Statistical Papers*, **55**(1), 141–156.
- Nordhausen K, Ollila E, Oja H (2011). “On the Performance Indices of ICA and Blind Source Separation.” In *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, pp. 486–490. IEEE.
- Papadimitriou C, Steiglitz K (1982). *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, Englewood Cliffs.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Risk BB, Matteson DS, Ruppert D, Eloyan A, Caffo BS (2014). “An Evaluation of Independent Component Analyses with an Application to Resting-State fMRI.” *Biometrics*, **70**(1), 224–236.
- Tan WY (1977). “On the Distribution of Quadratic Forms in Normal Random Variables.” *Canadian Journal of Statistics*, **5**(2), 241–250.
- Virta J, Li B, Nordhausen K, Oja H (2017). “Independent Component Analysis for Tensor-Valued Data.” *Journal of Multivariate Analysis*, **162**, 172–192.
- Virta J, Nordhausen K, Oja H (2016). “Projection Pursuit for Non-Gaussian Independent Components.” *arXiv preprint arXiv:1612.05445*.

Affiliation:

Pauliina Ilmonen
Department of Mathematics and Systems Analysis
Aalto University School of Science
P.O. Box 11100, FI-00076 Aalto
E-mail: pauliina.ilmonen@aalto.fi