MINIMUM DISTANCE INDEX FOR NON-SQUARE COMPLEX VALUED MIXING MATRICES

N. Lietzén[1], J. Virta[1], K. Nordhausen[2], P. Ilmonen[1]
[1]*Aalto University School of Science*, e-mail: pauliina.ilmonen@aalto.fi
[2]*Institute of Statistics & Mathematical Methods in Economics, TU Wien*
Helsinki, FINLAND; Vienna, AUSTRIA

Abstract

We consider complex valued linear blind source separation, where the objective is to estimate a subset of the latent sources. 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 square real valued (original) case.
Keywords: blind source separation, latent source, minimum distance index, data science

1 Introduction

In classic linear blind source separation (BSS), one assumes that the observation vectors are linear mixtures of a collection of latent *source* variables. The linearity assumption strikes a good balance between intricacy, mathematical tractability and interpretability, and the model has been used successfully in a wide variety of contexts ranging from signal processing and economics to biomedical applications. See, e.g., [1] for an introduction.

Numerous algorithms for the linear BSS problem under various assumptions on the sources have been proposed over the years, and a natural question is how to compare them. As the theoretical properties of the methods are often difficult to derive, comparisons are often conducted using simulation studies, see for example [4, 7, 8, 10, 12, 14]. To enable the comparisons, a performance measure for the success of the methods, called hereafter an *index*, is required. 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 [9] for a comparative study. MDI has the advantage of both, being affine invariant and having a known asymptotic behavior, in the case of real valued square mixing matrices, see [3, 5]. Therefore, in this paper, our focus is on the minimum distance index.

In this paper, we extend the MDI to the case where only some of the sources, the signals, are of interest to us, and the index should measure 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 work under the assumption of complex valued variables. A real valued version of this extension was proposed already in [13]. However, in contrast to the current paper, no theoretical justification for its properties was given in [13]. Similarly, a complex valued version of the regular MDI (where no distinction between the signal and the noise was made) was introduced in [6]. Moreover, the use of the minimum distance index is complicated by the fact that there is no clear scale attached to its values, making interpretation difficult. We provide a connection between the MDI and two commonly used statistics, SNR and correlation. These interpretations are given in the complex valued case, and the findings are novel also in the real valued case.

2 Complex valued blind source separation

Let $\mathbf{x}$ be an observable $\mathbb{C}^p$-valued random vector that obeys the complex valued blind source separation model,

$$\mathbf{x} = \boldsymbol{\Omega}\mathbf{z} = \boldsymbol{\Omega}\begin{pmatrix}\mathbf{z}_1^\top & \mathbf{z}_0^\top\end{pmatrix}^\top, \tag{1}$$

where $\boldsymbol{\Omega}$ 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 model, the $\mathbb{C}^d$-vector $\mathbf{z}_1$ contains the signals of interest that we wish to extract and the $\mathbb{C}^{d_0}$-vector $\mathbf{z}_0$, $d + d_0 = p$, contains uninteresting noise. Depending on the type of the problem, additional assumptions are imposed on the signal and the noise. For example, independent component analysis (ICA) assumes that $\mathbf{z}_1$ has independent marginals, second order separation (SOS) assumes that the signals and the noise have some specific temporal structure. In several approaches, some assumptions regarding the existence of the moments of $\mathbf{z}$ are also made.

The objective in Model (1) is to find a $\mathbb{C}^{d\times p}$ transformation matrix $\boldsymbol{\Gamma}$, such that the transformed vector corresponds to the signal variables, $\boldsymbol{\Gamma}\mathbf{x} = \mathbf{z}_1$, up to some class of transformations. The goal is to simultaneously extract both, the transformation matrix $\boldsymbol{\Gamma}$ and the source signals, by using only the information contained in the observable $\mathbf{x}$. Note that usually the transformation matrix $\boldsymbol{\Gamma}$ is not unique. For example, in ICA, the independent components stay independent, if we apply heterogeneous scaling, permutations or so-called phase-shifts to them. 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 the methods' performances. The minimum distance index discussed in the next section solves the issue by measuring how close the *gain matrix* $\mathbf{G} = \boldsymbol{\Gamma}\boldsymbol{\Omega}$ is to the matrix $\mathfrak{I}_{d,p} = \begin{pmatrix}\mathbf{I}_d & \mathbf{0}\end{pmatrix} \in \mathbb{R}^{d\times p}$, up to heterogeneous scaling, permutations and phase-shifts.

3 Non-square complex valued MDI

In this section, we provide the mathematical background for the non-square complex valued MDI. We begin with the concept of equivalent matrices in the sense that is relevant in the context of BSS.

Definition 1. Let $\sim$ be a relation on $\mathbb{C}^{d\times p}$, defined by $\mathbf{A} \sim \mathbf{B} \iff \mathbf{A} = (\mathbf{PJD})\mathbf{B}$, 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 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 $\exp(\theta_1 i),\ldots,\exp(\theta_d i)$, where i is the imaginary unit. Furthermore, we use the notation $\mathcal{C}^d$ for the set defined as $\mathcal{C}^d = \{\mathbf{C}\in\mathbb{C}^{d\times d} \mid \mathbf{C} = \mathbf{PJD} : \mathbf{P}\in\mathcal{P}^d, \mathbf{J}\in\mathcal{J}^d, \mathbf{D}\in\mathcal{D}^d\}$.

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

The proof of Proposition 2 is omitted here as reflexivity, symmetry and transitivity of the relation $\sim$ can be verified straightforwardly. We use the relation $\sim$ to partition $\mathbb{C}^{d\times p}$ into equivalence classes and denote 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 $\mathfrak{I}_{d,p} = \begin{pmatrix}\mathbf{I}_d & \mathbf{0}\end{pmatrix} \in\mathbb{R}^{d\times p}$. The optimization problem can be formulated as,

$$\text{minimize} \quad \|\mathbf{PJDA} - \mathfrak{I}_{d,p}\|_{\mathrm{F}}^2 \quad \text{s.t.} \quad \mathbf{P}\in\mathcal{P}^d, \quad \mathbf{J}\in\mathcal{J}^d \quad \text{and} \quad \mathbf{D}\in\mathcal{D}^d, \tag{2}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. Note that this optimization problem is not solvable in general, since the diagonal elements of matrices belonging to $\mathcal{D}^d$ are in the open interval $(0,\infty)$. Hereby, we define the shortest squared distance between the equivalence class $\mathcal{C}_{\mathbf{A}}$ and the matrix $\mathfrak{I}_{d,p}$ to be the greatest lower bound, that is,

$$\varrho\left(\mathbf{A},\mathfrak{I}_{d,p}\right) = \inf_{\mathbf{B}\in\mathcal{C}_{\mathbf{A}}} \|\mathbf{B} - \mathfrak{I}_{d,p}\|_{\mathrm{F}}^2 = \inf_{\mathbf{C}\in\mathcal{C}^d} \|\mathbf{CA} - \mathfrak{I}_{d,p}\|_{\mathrm{F}}^2 . \tag{3}$$

The distance $\varrho\left(\mathbf{A},\mathfrak{I}_{d,p}\right)$ defined above can be converted into a more applicable form from the computational point of view, see the following theorem.

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

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

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 also zeros on the diagonal. Then, 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. Now, since the Frobenius norm is orthogonally invariant, the objective function f can be reformulated as follows,

$$f\left(\mathbf{P},\mathbf{L},\mathfrak{I}_{d,p}\right) = \|\mathbf{PLA} - \mathfrak{I}_{d,p}\|_{\mathrm{F}}^2 = \left\|\mathbf{P}\left(\mathbf{LA} - \mathbf{P}^\top\mathfrak{I}_{d,p}\right)\right\|_{\mathrm{F}}^2 = \left\|\mathbf{LA} - \mathbf{P}^\top\mathfrak{I}_{d,p}\right\|_{\mathrm{F}}^2 .$$

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 elements. Now, since $\mathbf{L}$ is a diagonal matrix, we obtain the following form for $f(\mathbf{P},\mathbf{L},\mathfrak{I}_{d,p})$:

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

In the above formulation, 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 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. Regardless of the optimal permutation matrix, the partial derivatives with respect to the first ℓ diagonal elements of $\mathbf{Q}$ and $\mathbf{R}$ are,

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

from which we can solve the solution candidates $\mathbf{Q}'_{jj}$ and $\mathbf{R}'_{jj}$,

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

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,

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

such that $\pi : \mathcal{P}^d \times \{1,\ldots,d\} \to \{1,\ldots,d\} : \mathbf{P}\times j \mapsto c_j$, 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,

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

□

We next present the minimum distance index (MDI) for non-square complex valued mixing matrices. Note that this is an extension to the cases presented in [3, 5, 6].

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

$$\mathrm{MD}(\hat{\mathbf{\Gamma}}) = \frac{\varrho\left(\hat{\mathbf{G}},\mathfrak{I}_{d,p}\right)}{\sqrt{d}} = \frac{1}{\sqrt{d}}\inf_{\mathbf{C}\in\mathcal{C}^d}\left\|\mathbf{C}\hat{\mathbf{G}} - \begin{pmatrix}\mathbf{I}_d & \mathbf{0}\end{pmatrix}\right\|_{\mathrm{F}}.$$

Remark 5. In previous formulations of the MDI in [3, 5, 6], the MDI was defined such that ϱ was scaled by $1/\sqrt{d-1}$. In the corresponding previous work, it was assumed that every row of $\hat{\mathbf{G}}$ has at least one nonzero element. In this paper, we consider the more general situation allowing zero rows, yielding the scaling factor $1/\sqrt{d}$.

As in [5], 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 [11]. We next establish some key properties of the proposed index.

Theorem 6. *Let* $\boldsymbol{A} \in \mathbb{C}^{d\times p}$*. Then,* $\mathrm{MD}(\boldsymbol{A}) \in [0,1]$ *and the MDI satisfies,*

(i) $\mathrm{MD}(\boldsymbol{A}) = 0 \Longleftrightarrow \boldsymbol{A} \sim \begin{pmatrix}\boldsymbol{I}_d & \boldsymbol{0}\end{pmatrix} = \mathfrak{I}_{d,p}$*,*

(ii) $\mathrm{MD}(\boldsymbol{A}) = 1 \Longleftrightarrow \exists \boldsymbol{B} \in \mathbb{C}^{d\times(p-d)} : \boldsymbol{A} = \begin{pmatrix}\boldsymbol{0} & \boldsymbol{B}\end{pmatrix}$*,*

(iii) the function $f : c \mapsto \mathrm{MD}\left[\begin{pmatrix}\boldsymbol{I}_d & \boldsymbol{0}\end{pmatrix} + c\cdot \mathrm{off}(\boldsymbol{A})\right]$ *is increasing in* $c \in [0,1]$ *for all matrices* $\boldsymbol{A}$ *that satisfy* $|\boldsymbol{A}_{jk}| < 1$ *when* $j \neq k$*. (The function* $\mathrm{off}(\cdot)$ *here sets the main-diagonal elements of its argument equal to zero.)*

Proof of Theorem 6. By Theorem 3, we have that $\mathrm{MD}^2(\mathbf{A}) = 1 - \frac{1}{d}\max_{\mathbf{P}}\{\mathrm{Trace}(\mathbf{P}\tilde{\mathbf{A}})\}$, where the elements of $\tilde{\mathbf{A}}$ are between zero and one. It follows from this formulation that $\mathrm{MD}(\mathbf{A}) \in [0,1]$. Next, we proceed to verify properties (i)–(iii).

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

For property (ii), first assume that $\mathbf{A} = \begin{pmatrix}\mathbf{0} & \mathbf{B}\end{pmatrix}$. Then, the main-diagonal of $\tilde{\mathbf{A}}_{jk}$ contains only zeros, regardless of the permutation and regardless of the matrix $\mathbf{B}$. Thus, $\max_{\mathbf{P}}\{\mathrm{Trace}(\mathbf{P}\tilde{\mathbf{A}})\} = 0$, which gives us that $\mathrm{MD}(\mathbf{A}) = 1$. For the only if part, assume that $\mathrm{MD}(\mathbf{A}) = 1$. This assumption gives us that $\mathrm{Trace}(\mathbf{P}\tilde{\mathbf{A}}) = 0$ for the optimal $\mathbf{P}$ and consequently for every other $\mathbf{P}$ as well. Thus, $\mathbf{A} = \begin{pmatrix}\mathbf{0} & \mathbf{B}\end{pmatrix}$, 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,

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

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

4 Interpretation of the minimum distance index

The interpretation of the MDI presented in [3, 5] has been difficult, since its formulation provided no clear way of directly interpreting its numerical values. While the extreme values of 0 and 1 are by Theorem 6, respectively, indicators of a perfect and fully unsuccessful separation, the intermediate values have been complicated to understand.

We next provide connections between the MDI and two easily interpretable statistics, signal-to-noise ratio and correlation. Assume Model (1), where $\mathrm{Cov}(\mathbf{z}) = \mathbf{I}_p$ (taken without loss of generality as the scales of the sources are confounded with the scales of $\boldsymbol{\Omega}$). Assume furthermore that our aim is to extract only the d signal sources $\mathbf{z}_1 = (z_1, \ldots, z_d)$ and for this we have obtained an estimate $\hat{\boldsymbol{\Gamma}} \in \mathbb{C}^{d \times p}$, with the corresponding gain matrix $\hat{\mathbf{G}} = \hat{\boldsymbol{\Gamma}}\boldsymbol{\Omega} \in \mathbb{C}^{d \times p}$.

The estimates of the signals are then equal to $\hat{\mathbf{z}}_1 = (\hat{z}_1, \ldots, \hat{z}_d)^\top = \hat{\mathbf{G}}\mathbf{z}$ and we define the signal-to-noise ratio (SNR) and the correlation of the jth signal to be,

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

We here use the version of SNR defined in [2]. The following theorem connects the three statistics (MDI, SNR and correlation) under two different models for the gain matrix $\hat{\mathbf{G}}$: homogeneous contamination where all of the elements of $\hat{\mathbf{G}}$ deviate equally from those of $\mathfrak{I}_{d,p}$, and heterogeneous contamination where no structure for the contamination is assumed besides the requirement that it applies only to the off-diagonal elements of $\hat{\mathbf{G}}$. Let $\mathtt{A}(c_j) = \mathtt{A}(c_1, \ldots, c_d)$ and $\mathtt{H}(c_j) = \mathtt{H}(c_1, \ldots, c_d)$ denote, respectively, the arithmetic and harmonic means of the values $c_1, \ldots, c_d$.

Theorem 7. *Assume Model* (1)*, where* $\mathrm{Cov}(\boldsymbol{z}) = \boldsymbol{I}_p$*. Then, under the homogeneous contamination model* $\hat{\boldsymbol{G}} = \mathfrak{I}_{d,p} + \varepsilon \boldsymbol{1}_{d,p}$*, where* $\boldsymbol{1}_{d,p}$ *is a* $d \times p$ *matrix full of ones and* $\mathrm{Re}(\varepsilon) > -1/2$*,*

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

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

$$\mathrm{MD}^2(\hat{\boldsymbol{G}}) = \mathtt{H}(\mathrm{SNR}_1, \ldots, \mathrm{SNR}_d)^{-1} = 1 - \mathtt{A}(|\mathrm{Cor}_1|^2, \ldots, |\mathrm{Cor}_d|^2).$$

Proof of Theorem 7. We consider the two models separately. Under the homogeneous one, we have $\mathrm{MD}^2(\hat{\mathbf{G}}) = 1 - \frac{1}{d}\max_{\mathbf{P}}\{\mathrm{Trace}(\mathbf{P}\tilde{\mathbf{G}})\}$, where the diagonal elements of $\tilde{\mathbf{G}}$ are equal to $|1+\varepsilon|^2/[|1+\varepsilon|^2 + (p-1)|\varepsilon|^2]$ and the off-diagonal elements are equal to $|\varepsilon|^2/[|1+\varepsilon|^2 + (p-1)|\varepsilon|^2]$. The requirement that $\mathrm{Re}(\varepsilon) > -1/2$ ensures that the maximization over $d \times d$ permutation matrices is solved by $\mathbf{I}_d$ and we get $\mathrm{MD}^2(\hat{\mathbf{G}}) = [(p-1)|\varepsilon|^2]/[|1+\varepsilon|^2 + (p-1)|\varepsilon|^2]$. Consider then the signal-to-noise ratio SNR_j of the jth signal $\hat{z}_j = \hat{\mathbf{G}}_{j1}z_1 + \cdots + \hat{\mathbf{G}}_{jd}z_p$. As the sources are uncorrelated and have unit variances, we obtain

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

Table 1: Average SNRs and correlations corresponding to certain MDI-values.

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

Similarly, for the correlation, we have $\text{Var}(z_j) = 1$, $\text{Var}(\hat{z}_j) = |1+\varepsilon|^2 + (p-1)\,|\varepsilon|^2$ and $\text{Cov}(\hat{z}_j, z_j) = 1 + \varepsilon$, yielding,

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

proving the claim under the homogeneous contamination. The proof for the heterogeneous contamination proceeds in exactly the same manner and we provide only the values for the key quantities. The squared MDI equals $\text{MD}^2(\hat{\mathbf{G}}) = 1 - (1/d)\sum_{j=1}^{d}(1 + \sum_{k\neq j}^{p} |\hat{\mathbf{G}}_{jk}|^2)^{-1}$. Using the general form for signal-to-noise ratio in (4), we get $1 - (\text{SNR}_j)^{-1} = (1 + \sum_{k\neq j}^{p} |\hat{\mathbf{G}}_{jk}|^2)^{-1}$ which when combined with the expression for $\text{MD}^2(\hat{\mathbf{G}})$ yields the claim for SNR. For the correlation we have $\text{Var}(z_j) = 1$, $\text{Var}(\hat{z}_j) = 1 + \sum_{k\neq j}^{p} |\hat{\mathbf{G}}_{jk}|^2$ and $\text{Cov}(\hat{z}_j, z_j) = 1$, yielding $|\text{Cor}_j|^2 = (1 + \sum_{k\neq j}^{p} |\hat{\mathbf{G}}_{jk}|^2)^{-1}$ and establishing the final part of the claim. □

Table 1 displays the connection between the MDI values and the average signal-to-noise ratios and the correlations under the heterogeneous contamination model. In the table, the signal-to-noise ratios have been converted to the standard decibel scale using the transformation $10\log_{10}(\text{SNR})$. Interestingly, the table shows that the minimum distance index of 0.1 can already be considered extremely good; it corresponds to the high average SNR of 20 dB and to the almost perfect average correlation of 0.995. Similarly, the index value of 0.01 can be considered to correspond to an almost flawless separation.

5 Discussion

In recent years BSS has become a popular tool for dimension reduction and thus the interest for models with non-square mixing matrices has grown considerably. To facilitate comparisons of different BSS methods, we extended the popular MDI to the complex valued non-square case. To give a better understanding of the MDI values besides the extremes, we derived a connection between MDI, and SNR and correlations, which provides the opportunity to better judge if a method is of practical use.

In future work, we will derive asymptotic properties of the extended MDI. Moreover, we will explore whether variants of the index can be used in nonlinear BSS.

Acknowledgements: N. Lietzén gratefully acknowledges financial support from the Emil Aaltonen Foundation (grant 180144). The authors would also like to thank the anonymous referee.

References

[1] Comon P., Jutten C. (2010). *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic Press, Oxford.

[2] Gonzales R.C., Woods R.E. (2001). *Digital Image Processing, Second Edition*. Prentice Hall, Upper Saddle River, New Jersey.

[3] Ilmonen P., Nordhausen K., Oja H., Ollila E. (2010). A new performance index for ICA: properties, computation and asymptotic analysis. *International Conference on Latent Variable Analysis and Signal Separation*. pp. 229-236, Springer.

[4] Ilmonen P., Paindaveine D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Annals of Statistics*. Vol. **39**, pp. 2448-2476.

[5] Ilmonen P., Nordhausen K., Oja H., Ollila E. (2012). On asymptotics of ICA estimators and their performance indices. *arXiv preprint arXiv:1212.3953*.

[6] Lietzén N., Nordhausen K., Ilmonen P. (2016). Minimum distance index for complex valued ICA. *Statistics & Probability Letters*. Vol. **118**, pp. 100-106.

[7] Matteson D.S., Tsay R.S. (2017). Independent component analysis via distance covariance. *Journal of the American Statistical Association*. Vol. **112**, pp. 623-637.

[8] Miettinen J., Illner K., Nordhausen K., Oja H., Taskinen S.,Theis F.J. (2016). Separation of uncorrelated stationary time series using autocovariance matrices. *Journal of Time Series Analysis*. Vol. **37**, pp. 337-354.

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

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

[11] Papadimitriou C., Steiglitz K. (1982). *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, Englewood Cliffs.

[12] Risk B.B., Matteson D.S., Ruppert D., Eloyan A., Caffo B.S. (2014). An evaluation of independent component analyses with an application to resting-state fMRI. *Biometrics*. Vol. **70**, pp. 224-236.

[13] Virta J., Nordhausen K., Oja H. (2016). Projection pursuit for non-Gaussian independent components. *arXiv preprint arXiv:1612.05445*.

[14] Virta J., Li B., Nordhausen K., Oja H. (2017). Independent component analysis for tensor-valued data. *Journal of Multivariate Analysis*, Vol. **162**, pp. 172-192.