Measure-Transformed SOBI

Sapir Shem Tov, Koby Todros and Igal Bilik
Ben-Gurion University of the Negev

***Abstract*—This paper addresses the problem of robust blind source separation (BSS) in the presence of non-Gaussian heavy-tailed noise. We focus on overdetermined linear mixtures of temporally correlated stationary signals, which are corrupted by additive stationary noise. Within this framework, we introduce a new separation method called measure-transformed second-order blind identification (MT-SOBI). This method operates by transforming pairwise joint probability measures of time-lagged samples. The considered transform is generated by a non-negative data-weighting function, referred to as MT-function. We demonstrate that proper selection of the MT-function can result in significant resilience to heavy-tailed noise while preserving the spatial correlation structure essential to perform separation. The MT-SOBI is illustrated in a simulation study highlighting its advantages over the standard SOBI and other robust extensions.**

***Index Terms*—Blind source separation, parameter estimation, probability measure transform, robust statistics.**

I. Introduction

Blind source separation (BSS) is a fundamental problem in statistical signal processing, aiming to recover latent source signals from an observed mixture without prior knowledge of the source distributions or the mixing process [1]. BSS plays an important role in speech enhancement, biomedical signal processing, communications, and image processing.

The second-order blind identification (SOBI) algorithm [2] is a well-known method for separating linear mixtures of spatially uncorrelated sources by leveraging temporal correlations. The conventional SOBI is formulated under a noisy observation model, where the noise is additive and spatially white. In SOBI, the data undergoes pre-whitening using a whitening matrix derived from the sample covariance matrix. Source separation is then performed via orthogonal approximate joint diagonalization of sample autocorrelation matrices at different time-lags, utilizing successive Givens rotations.

However, the sample covariance and autocorrelation estimators are highly sensitive to heavy-tailed noise, which produces outliers. Under such a scenario, these estimators provide inaccurate representations of the underlying covariance and autocorrelation structures, leading to compromised separation performance.

To address this limitation, several robust SOBI variants have been proposed. In [3] and [4], the SAM-SOBI and eSAM-SOBI algorithms were introduced, respectively, wherein the sample autocorrelation matrices are replaced by robust empirical sign-autocorrelation matrices. The key distinction between the two lies in the estimation of the whitening matrix: the former employs an empirical sign-covariance matrix [5], while the latter applies the affine equivariant Hettmansperger-Randles covariance estimator [6] instead. It is important to note that both SAM-SOBI and eSAM-SOBI are formulated under the assumption of a noiseless mixing model, making them more suitable when data contamination follows Huber's ϵ-contamination model [7] rather than when affected by additive heavy-tailed noise. In the presence of additive noise, the sign-autocorrelation matrix does not necessarily preserve the spatial correlation structure, which is essential for consistent estimation of the mixing matrix. Additionally, the influence function [8] of the empirical sign-autocorrelation matrix, which quantifies its sensitivity to outliers, is bounded, meaning that this estimator is B-robust [8]. However, as demonstrated in [9, Sec. III-A] for the empirical sign-covariance matrix, the influence function does not vanish as the outlier's magnitude increases. This implies that while the estimator is resilient against a small amount of contamination, it does not reject large-norm outliers.

Another robust SOBI adaptation, called MF-SOBI [10], incorporates median filtering as a preprocessing step to mitigate outliers. However, it is important to note that non-linear median filtering alters the linear mixing structure, potentially leading to inconsistent estimation of the mixing matrix. Lastly, [11] introduced a robust recursive SOBI extension designed for non-stationary environments.

Main contribution: In this paper, we introduce a new robust SOBI variant tailored for overdetermined linear mixtures of temporally correlated wide-sense stationary (WSS) sources corrupted by additive stationary heavy-tailed noise. Specifically, the noise distribution is assumed to belong to the broad class of compound Gaussian (CG) distributions [12], which encompasses both heavy-tailed distributions (e.g., the t-distribution) and light-tailed ones (e.g., Gaussian).

The proposed method, called measure-transformed (MT)-SOBI operates by applying a transform to pairwise joint probability measures of time-lagged samples. The considered transform [9], [13]-[15] is generated by a non-negative data-weighting function, called MT-function. Proper selection of the MT-function can lead to suppression of zones in the observation space attributed to outliers, leading to significantly enhanced separation performance.

The MT-SOBI consists of two steps. First, a consistent estimate of the projection matrix onto the null space of the transposed mixing matrix is obtained through spectral decomposition of an empirical MT zero-lagged autocorrelation matrix. In the second step, this projection matrix is utilized to construct a set of empirical MT-autocorrelation matrices corresponding to distinct non-zero time-lags. A consistent estimate of the non-square mixing matrix is then derived via non-orthogonal approximate joint diagonalization (AJD) [16]

This work was supported in part by Israel Innovation Authority under grant 84001.

of these matrices.

In this context, we note that the measure transformation framework has been previously applied to BSS in [14]. However, unlike MT-SOBI, which leverages temporal correlations at different time-lags within a single WSS segment, the method in [14] is tailored for piecewise-stationary mixtures and exploits variance diversity across independent data segments.

The MT-SOBI offers several advantages over the aforementioned techniques. First, unlike SOBI and its robust variants, MT-SOBI does not apply pre-whitening, thereby retaining the original data structure. Furthermore, in contrast to the sign-autocorrelation matrix used in SAM-SOBI and eSAM-SOBI, the MT-autocorrelation matrix preserves the spatial correlation structure even in the presence of additive noise, which is essential for consistent mixing matrix estimation. Moreover, when employing specific Gaussian MT-functions, the empirical MT-autocorrelation matrix exhibits B-robustness, with an influence function that decays to zero with the outlier norm. Consequently, unlike SAM-SOBI and eSAM-SOBI, the MT-SOBI effectively rejects large-norm outliers, leading to more effective outlier suppression. These advantages translate into significant performance improvements, as demonstrated by simulations under both light-tailed Gaussian and heavy-tailed non-Gaussian noise conditions.

II. Probability measure transform: Review

In this section, we present an overview of some fundamental principles of the probability measure transformation framework [9], [13]-[15]. These principles will be subsequently utilized to develop the proposed MT-SOBI algorithm.

A. Probability measure transform

Let $\mathbf{x}$ and $\mathbf{y}$ represent random vectors, with their respective observation spaces denoted as $\mathcal{X} \subseteq \mathbb{R}^p$ and $\mathcal{Y} \subseteq \mathbb{R}^q$. We define the measure space $(\mathcal{X} \times \mathcal{Y}, \mathcal{S}_{\mathcal{X}\times\mathcal{Y}}, P_{\mathbf{xy}})$, where $\mathcal{S}_{\mathcal{X}\times\mathcal{Y}}$ is a σ -algebra over the Cartesian product $\mathcal{X} \times \mathcal{Y}$ and $P_{\mathbf{xy}}$ is the joint probability measure defined on $\mathcal{S}_{\mathcal{X}\times\mathcal{Y}}$.

Definition 1. *Let $u : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}_+$ represent a non-negative function, such that the respective statistical expectation $\mathbb{E}[u(\mathbf{x},\mathbf{y}); P_{\mathbf{xy}}] \triangleq \int_{\mathcal{X}\times\mathcal{Y}} u(\mathbf{r},\boldsymbol{t})\, dP_{\mathbf{xy}}(\mathbf{r},\boldsymbol{t})$ is finite and strictly positive. A transform on the probability measure $P_{\mathbf{xy}}$ is defined as:*

$$Q^{(u)}_{\mathbf{xy}}(\mathcal{A}) \triangleq \mathbb{T}_u[P_{\mathbf{xy}}](\mathcal{A}) = \int_{\mathcal{A}} \varphi_u(\mathbf{r},\boldsymbol{t})\, dP_{\mathbf{xy}}(\mathbf{r},\boldsymbol{t}), \quad (1)$$

where $\mathcal{A} \in \mathcal{S}_{\mathcal{X}\times\mathcal{Y}}$ and $\varphi_u(\mathbf{r},\boldsymbol{t}) \triangleq u(\mathbf{r},\boldsymbol{t})/\mathbb{E}[u(\mathbf{x},\mathbf{y}); P_{\mathbf{xy}}]$. The function $u(\cdot,\cdot)$ is the generating function of the transform and will be referred to as the MT-function.

Definition 1 implies that $Q^{(u)}_{\mathbf{xy}}$ is a probability measure on $\mathcal{S}_{\mathcal{X}\times\mathcal{Y}}$ that is absolutely continuous with respect to $P_{\mathbf{xy}}$. The corresponding Radon-Nikodym derivative [17] is given by:

$$dQ^{(u)}_{\mathbf{xy}}(\mathbf{r},\boldsymbol{t})/dP_{\mathbf{xy}}(\mathbf{r},\boldsymbol{t}) \triangleq \varphi_u(\mathbf{r},\boldsymbol{t}). \quad (2)$$

In practice, applying the transform to a probability distribution $P_{\mathbf{xy}}$ imposes weighting to the data samples drawn from it. This weighting is controlled by the MT-function $u(\cdot,\cdot)$, which, when appropriately chosen, can effectively mitigate the influence of noisy outliers.

B. The measure-transformed cross-correlation

Relation (2) implies that the cross-correlation matrix of $\mathbf{x}$ and $\mathbf{y}$ under the transformed probability measure $Q^{(u)}_{\mathbf{xy}}$ takes the form:

$$\mathbf{R}^{(u)}_{\mathbf{xy}} \triangleq \mathbb{E}[\mathbf{x}\mathbf{y}^T; Q^{(u)}_{\mathbf{xY}}] = \mathbb{E}[\mathbf{x}\mathbf{y}^T \varphi_u(\mathbf{x},\mathbf{y}); P_{\mathbf{xy}}]. \quad (3)$$

The matrix $\mathbf{R}^{(u)}_{\mathbf{xy}}$ is referred to as the MT-cross-correlation matrix. It represents a weighted cross-correlation under $P_{\mathbf{xy}}$, where the weighting function $\varphi_u(\cdot,\cdot)$ is defined below (1). Moreover, when the MT-function $u(\cdot,\cdot)$ is non-zero and constant valued, the transformed probability measure $Q^{(u)}_{\mathbf{xY}}$ coincides with the origin measure $P_{\mathbf{xy}}$. In this case, the MT-cross-correlation reduces to the conventional cross-correlation matrix of two random vectors.

Given a sequence of K pairwise samples $\{(\mathbf{x}_n,\mathbf{y}_n)\}_{n=1}^K$ generated from $P_{\mathbf{xy}}$, the empirical MT-cross-correlation matrix is defined as:

$$\hat{\mathbf{R}}^{(u)}_{\mathbf{xy}} \triangleq \frac{1}{K}\sum\nolimits_{n=1}^{K} \mathbf{x}_n\mathbf{y}_n^T \hat{\varphi}_u(\mathbf{x}_n,\mathbf{y}_n), \quad (4)$$

where $\hat{\varphi}_u(\mathbf{r},\boldsymbol{t}) \triangleq u(\mathbf{r},\boldsymbol{t})/(K^{-1}\sum_{n=1}^K u(\mathbf{x}_n,\mathbf{y}_n))$. The infinitesimal robustness of this estimator against outliers is evaluated using its influence function [8]. This function quantifies the effect on the asymptotic bias caused by an infinitesimal amount of contamination at some point in the joint observation space, which represents an outlier. Similarly to the influence function of the empirical MT-covariance [9], it can be shown that the influence function associated with the empirical MT-cross-correlation matrix (4) takes the form:

$$\mathbf{IF}(\mathbf{r},\boldsymbol{t}) = \frac{u(\mathbf{r},\boldsymbol{t})(\mathbf{r}\boldsymbol{t}^T - \mathbf{R}^{(u)}_{\mathbf{xy}})}{\mathbb{E}[u(\mathbf{x},\mathbf{y}); P_{\mathbf{xy}}]}, \quad (5)$$

where the pair $(\mathbf{r},\boldsymbol{t}) \in \mathcal{X}\times\mathcal{Y}$ represents an outlying point in the joint observation space. An estimator is considered B-robust if its influence function is bounded [8]. For the MT-cross-correlation estimator, it follows from (5) that this property holds if the MT-function $u(\cdot,\cdot)$ satisfies

$$u(\mathbf{r},\boldsymbol{t}) \le C \;\text{ and }\; u(\mathbf{r},\boldsymbol{t})(\|\mathbf{r}\|^2 + \|\boldsymbol{t}\|^2) \le C, \quad (6)$$

where C is a positive constant. Additionally, an estimator is said to reject large-norm outliers if its influence function decays to zero as the outlier norm approaches infinity. Based on (5), this property holds if the MT-function $u(\cdot,\cdot)$ satisfies

$$u(\mathbf{r},\boldsymbol{t})(\|\mathbf{r}\|^2 + \|\boldsymbol{t}\|^2) \to 0 \;\text{ as }\; \|\mathbf{r}\|^2 + \|\boldsymbol{t}\|^2 \to \infty. \quad (7)$$

III. Problem formulation

Consider an observation sequence that conforms to a noisy linear mixing model:

$$\mathbf{x}_n = \mathbf{A}\boldsymbol{s}_n + \mathbf{w}_n, \quad n = 1,\ldots,N, \quad (8)$$

where $\{\mathbf{x}_n \in \mathbb{R}^p\}$ is an observation process and n denotes a discrete time instance. The processes $\{\boldsymbol{s}_n \in \mathbb{R}^q, q < p\}$

and $\{\mathbf{w}_n \in \mathbb{R}^p\}$ represent latent source and noise processes respectively. The matrix $\mathbf{A} \in \mathbb{R}^{p\times q}$ is an unknown full-column rank mixing matrix. We assume that the source and noise processes satisfy the following conditions:

(A-1) The source process $\{\boldsymbol{s}_n\}$ is identically distributed, with symmetric distribution about the origin and a non-singular covariance matrix at each time instance n.

(A-2) The source process $\{\boldsymbol{s}_n\}$ is WSS. The respective spatial autocorrelation matrix $\mathbf{R}_{\boldsymbol{s}}(m) \triangleq \mathbb{E}[\boldsymbol{s}_n \boldsymbol{s}_{n+m}^T; P_{\boldsymbol{s}_n \boldsymbol{s}_{n+m}}]$ is diagonal for any time-lag m.

(A-3) The noise process $\{\mathbf{w}_n\}$ is independent and identically distributed (i.i.d.).

(A-4) The noise process $\{\mathbf{w}_n\}$ obeys a CG distribution [12] at each time instance, with stochastic representation:

$$\mathbf{w} = \sqrt{\nu}\boldsymbol{\psi}, \tag{9}$$

where ν is a non-negative random variable and $\boldsymbol{\psi} \in \mathbb{R}^p$ is a zero-mean Gaussian random vector, independent of ν, with scaled-identity covariance matrix $\sigma_\psi^2 \mathbf{I}_p$.

(A-5) $\{\boldsymbol{s}_n\}$ and $\{\mathbf{w}_n\}$ are statistically independent.

Given the data samples in (8), our goal is to estimate the mixing matrix $\mathbf{A}$ and recover the source signals. Once an estimate $\hat{\mathbf{A}}$ of the mixing matrix is obtained, the sources can be recovered by computing the separation matrix $\mathbf{B} = \hat{\mathbf{A}}^\dagger$, where $\hat{\mathbf{A}}^\dagger \triangleq (\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^T$ is the Moore-Penrose pseudo inverse [18] of $\hat{\mathbf{A}}$. The recovered sources are then given by $\{\hat{\boldsymbol{s}}_n = \mathbf{B}\mathbf{x}_n = \boldsymbol{s}_n + \mathbf{B}\mathbf{w}_n,\ \ n = 1, \ldots, N\}$.

IV. Blind identification

In this section, we introduce a two-step procedure for blind identification of the non-square mixing matrix $\mathbf{A}$, up to the inherent ambiguities of scaling and column permutations. This procedure forms the basis for the proposed MT-BSS algorithm outlined in the next section. The initial phase of this identification procedure extracts the projection matrix onto the null space of $\mathbf{A}^T$ from a zero-lagged MT-autocorrelation matrix. The subsequent step leverages this projection matrix to derive a set of MT-autocorrelation matrices at different time-lags, from which $\mathbf{A}$ can be identified via joint diagonalization.

A. Stage I: Extraction of the projection matrix onto $\mathrm{null}(\mathbf{A}^T)$

First, we note that the observation process $\{\mathbf{x}_n\}$ introduced in (8) is identically distributed. Hence, it follows from (3) that the zero-lagged MT-autocorrelation matrix $\mathbf{R}_{\mathbf{x}}^{(u)}(0) \triangleq \mathbf{R}_{\mathbf{x}_n\mathbf{x}_n}^{(u)}$ (which is the MT-cross-correlation of $\mathbf{x}_n$ and itself) must be time-invariant. Next, we consider the following MT-function:

$$u(\mathbf{r}, \boldsymbol{t}) \triangleq h(\|\mathbf{r}\|) \times h(\|\boldsymbol{t}\|), \tag{10}$$

where $h : \mathbb{R} \to \mathbb{R}$ is a non-negative zero-centered symmetric function. Under Assumptions (A-1) and (A-3)-(A-5), it follows from [15, Th. 2] that the respective time-invariant zero-lagged MT-autocorrelation matrix takes the form:

$$\mathbf{R}_{\mathbf{x}}^{(u)}(0) = \mathbf{A}\mathbf{B}_u\mathbf{A}^T + c_u\mathbf{I}_p, \tag{11}$$

where $\mathbf{B}_u \in \mathbb{R}^{q\times q}$ is a positive-definite matrix and c_u is a positive scalar. Since $\mathbf{A}$ is assumed to have a full column rank, this structure facilitates the determination of the null space of $\mathbf{A}^T$ through eigen-decomposition of $\mathbf{R}_{\mathbf{x}}^{(u)}(0)$. The null space is spanned by the $p-q$ eigenvectors corresponding to the smallest $p-q$ eigenvalues. Let $\mathbf{V}_u \in \mathbb{R}^{p\times(p-q)}$ denote the matrix comprised of these vectors. Then, the projection matrix onto $\mathrm{null}(\mathbf{A}^T)$ is obtained by:

$$\mathbf{P}_{\mathbf{A}}^{\perp} = \mathbf{V}_u\mathbf{V}_u^T. \tag{12}$$

B. Stage II: Identification of $\mathbf{A}$ via joint diagonalization

In the second step, we consider an MT-function of the form:

$$v(\mathbf{r}, \boldsymbol{t}) \triangleq g(\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{r}) \times g(\mathbf{P}_{\mathbf{A}}^{\perp}\boldsymbol{t}), \tag{13}$$

where $g : \mathbb{R}^p \to \mathbb{R}$ is a symmetric, zero-centered, and non-negative function. Using (13), we derive the corresponding MT-autocorrelation matrix of the observation process $\{\mathbf{x}_n\}$ (8) at an arbitrary time-lag m. This matrix is obtained from the MT-cross-correlation (3) by substituting $\mathbf{x}$ and $\mathbf{y}$ with $\mathbf{x}_n$ and $\mathbf{x}_{n+m}$, respectively. By (8) and (13), it follows that $v(\mathbf{x}_n, \mathbf{x}_{n+m}) = g(\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{w}_n) \times g(\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{w}_{n+m})$, implying that the projection matrix nullifies the dependency of the MT-function on the source vectors. Leveraging this key property along with Assumptions (A-1)-(A-3) and (A-5), we conclude that the MT-autocorrelation matrix is time-invariant and takes the form:

$$\mathbf{R}_{\mathbf{x}}^{(v)}(m) \triangleq \mathbf{R}_{\mathbf{x}_n,\mathbf{x}_{n+m}}^{(v)} = \mathbf{A}\mathbf{R}_{\boldsymbol{s}}(m)\mathbf{A}^T + \mathbf{R}_{\mathbf{w}}^{(v)}(0)\delta(m), \tag{14}$$

where $\mathbf{R}_{\boldsymbol{s}}(m)$ is the autocorrelation of the source process, defined in (A-2), $\mathbf{R}_{\mathbf{w}}^{(v)}(0)$ is the zero-lagged MT-autocorrelation of the noise process, and $\delta(\cdot)$ denotes Kronecker's delta.

The result in (14) implies that for any set $\mathcal{M}$ of *non-zero* time-lags, the corresponding MT-autocorrelation matrices are noise independent satisfying:

$$\left\{\mathbf{R}_{\mathbf{x}}^{(v)}(m) = \mathbf{A}\mathbf{R}_{\boldsymbol{s}}(m)\mathbf{A}^T\right\}_{m\in\mathcal{M}}, \tag{15}$$

where $\{\mathbf{R}_{\boldsymbol{s}}(m)\}_{m\in\mathcal{M}}$ are diagonal by assumption (A-2). Building on this outcome, the following proposition provides conditions under which the mixing matrix $\mathbf{A}$ can be recovered via joint diagonalization of $\{\mathbf{R}_{\mathbf{x}}^{(v)}(m)\}_{m\in\mathcal{M}}$.

Proposition 1 (Identification of the mixing matrix)**.** *Let $\mathcal{M}$ denote a set of distinct non-zero time-lags. Assume that the matrices $\{\mathbf{R}_{\boldsymbol{s}}(m)\}_{m\in\mathcal{M}}$ satisfy the following conditions:*

(B-1) Define $r_{i,m}$ as the i-th diagonal entry of $\mathbf{R}_{\boldsymbol{s}}(m)$. For each pair of distinct diagonal entries (i, j), there exists a pair of time-lag indices, (m, n), such that the vectors $[r_{i,m}, r_{i,n}]^T$ and $[r_{j,m}, r_{j,n}]^T$ are neither collinear nor equal to zero. This condition necessitates at least two distinct time-lags, i.e., $M \triangleq |\mathcal{M}| \geq 2$.

(B-2) There exists a time-lag $m^* \in \mathcal{M}$, such that the matrix $\mathbf{R}_s(m^*)$ is non-singular.

Under these conditions, the non-square full-column-rank mixing matrix $\mathbf{A}$ can be uniquely determined, up to scaling and column permutations, from the set of equations in (15).

The proof follows a similar approach to that of [14, Prop. 1]. However, it is noteworthy that in this formulation, the minimum required number M of jointly diagonalizable matrices is 2, whereas in [14, Prop. 1], it is 3.

V. The MT-SOBI algorithm

The MT-SOBI algorithm is derived by substituting the MT-autocorrelation matrices $\mathbf{R}_{\mathbf{x}}^{(u)}(0)$ and $\{\mathbf{R}_{\mathbf{x}}^{(v)}(m)\}_{m\in\mathcal{M}}$, employed in the two-step identification procedure outlined in Section IV, with their empirical estimates. At each stage, a scaled outlier-suppressive MT-function is chosen.

A. Stage I: Estimation of the projection matrix onto $\mathrm{null}(\mathbf{A}^T)$

At the initial phase, an empirical estimate of the projection matrix $\mathbf{P}_{\mathbf{A}}^{\perp}$ (12) is obtained as:

$$\hat{\mathbf{P}}_{\mathbf{A}}^{(u)\perp} \triangleq \hat{\mathbf{V}}_u \hat{\mathbf{V}}_u^T, \tag{16}$$

where $\hat{\mathbf{V}}_u \in \mathbb{R}^{p\times(p-q)}$ consists of the eigenvectors of the empirical zero-lagged MT-autocorrelation matrix $\hat{\mathbf{R}}_{\mathbf{x}}^{(u)}(0)$ associated with its $p-q$ smallest eigenvalues. The matrix $\hat{\mathbf{R}}_{\mathbf{x}}^{(u)}(0)$ is derived from (4) by substituting K with the actual sample size N and replacing $\mathbf{y}_n$ with $\mathbf{x}_n$ (8).

The MT-function (10) is specified within a parametric family consisting of pairwise products of scaled functions:

$$\{u(\mathbf{r},\boldsymbol{t};\theta) \triangleq h(\theta^{-1}\|\mathbf{r}\|)\times h(\theta^{-1}\|\boldsymbol{t}\|):\ \ \theta\in\mathbb{R}_{++}\}, \tag{17}$$

such that the robustness conditions (6) and (7) are satisfied. In this setting, we note that choosing a Gaussian shaped function, specifically $h(r)=\exp(-r^2)$, meets these conditions. The parameter θ is a strictly positive scaling factor that governs the suppression level of outliers. Here, θ is determined as a measure of nominal data dispersion:

$$\theta = 3\times p^{-1}\sum\nolimits_{i=1}^{p}\hat{\sigma}_i, \tag{18}$$

where $\hat{\sigma}_i \triangleq \kappa\times\mathrm{MAD}(\{x_{i,n}\}_{n=1}^N)$, $\kappa=1/\mathrm{erf}^{-1}(3/4)$, is a robust median absolute deviation (MAD) estimate of standard deviation [7] and $x_{i,n}$ is the i-th coordinate of the observation vector $\mathbf{x}_n$.

B. Stage II: Estimation of $\mathbf{A}$ via non-orthogonal AJD

In the second stage, the empirical projection matrix $\hat{\mathbf{P}}_{\mathbf{A}}^{(u)\perp}$ (16) is plugged into the MT-function $v(\cdot,\cdot)$ (13). Subsequently, based on Proposition 1, the non-square mixing matrix $\mathbf{A}$ is estimated by applying *non-orthogonal* AJD to the corresponding set of symmetrized empirical MT-autocorrelation matrices

$$\left\{\tilde{\mathbf{R}}_{\mathbf{x}}^{(v)}(m) \triangleq \frac{1}{2}\left(\hat{\mathbf{R}}_{\mathbf{x}}^{(v)}(m)+\hat{\mathbf{R}}_{\mathbf{x}}^{(v)T}(m)\right)\right\}_{m\in\mathcal{M}}. \tag{19}$$

Following the definition of the MT-autocorrelation matrix in (14), $\hat{\mathbf{R}}_{\mathbf{x}}^{(v)}(m)$ is derived from (4) by replacing K with $N-m$ and substituting $\mathbf{y}_n$ with $\mathbf{x}_{n+m}$. The averaging in (19) ensures symmetry in the empirical MT-autocorrelation matrices arising from (15). A viable choice for an AJD algorithm, suited for the given setting involving *non-square* mixing and *non-positive-definite* matrices, is the ACDC procedure described in [16].

Similarly to the previous stage, the MT-function (13) is specified within a parametric family comprising pairwise products of scaled functions:

$$\{v(\mathbf{r},\boldsymbol{t};\theta) \triangleq g(\theta^{-1}\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{r})\times g(\theta^{-1}\mathbf{P}_{\mathbf{A}}^{\perp}\boldsymbol{t}):\ \ \theta\in\mathbb{R}_{++}\}, \tag{20}$$

such that the robustness conditions (6) and (7) are satisfied. Analogously to the proof of [13, Prop. 7], it can be shown that under the choice of a Gaussian shaped function, specifically $g(\mathbf{r})=\exp(-\|\mathbf{r}\|^2)$, these conditions hold over a sufficiently large subspace of $\mathbb{R}^p\times\mathbb{R}^p$ whose probability measure is ≈ 1. Also here, the parameter θ is a strictly positive scaling factor controlling the shrinkage level of outliers. This parameter is chosen using the same selection rule outlined in (18).

VI. Numerical examples

In this section, we assess the performance of the proposed MT-SOBI in comparison to the non-robust SOBI [2] and the following robust alternatives: SAM-SOBI [3], eSAM-SOBI [4], and MF-SOBI [10]. To quantify estimation accuracy, we evaluate the deviation of $\hat{\mathbf{A}}^{\dagger}\mathbf{A}$ from the identity matrix using Amari's error [19, Sec. 5].

A. Experimental settings

Throughout this experiment, the dimensions of the observation and source processes were set to $p=10$ and $q=5$, respectively. The sample size was fixed at $N=1000$. The estimation accuracy was assessed over $J=10^3$ independent Monte-Carlo trials. In each trial, the entries of the mixing matrix $\mathbf{A}$ were drawn independently from a standard normal distribution, and each column was subsequently normalized to have unit Euclidean norm. Each source component sequence $\{s_{i,n}\}_{n=1}^N$, $i\in\{1,\ldots,q\}$, was modeled as a second-order autoregressive process, generated independently by passing a unit variance white Gaussian noise through the transfer function $A(z)=((1-\alpha z^{-1})(1-\alpha^* z^{-1}))^{-1}$. The pole $\alpha=\rho e^{j\phi}$ had magnitude ρ and phase ϕ, drawn independently from uniform distributions over the open intervals $(0.98, 0.99)$ and $(0.1\pi, 0.9\pi)$, respectively. We examined two types of the CG noise distributions: 1) light-tailed Gaussian distribution, and 2) heavy-tailed t-distribution with 3 degrees of freedom. The corresponding scatter parameter $\sigma_{\boldsymbol{\psi}}^2$, defined below (9), was chosen to maintain a fixed generalized signal-to-noise ratio (GSNR), defined here as $\mathrm{GSNR}\triangleq \mathrm{tr}[\mathbf{A}\mathbf{R}_{\boldsymbol{\mathcal{S}}}(0)\mathbf{A}^T]/(p\sigma_{\boldsymbol{\psi}}^2)$.

B. Implementation details

Throughout the study, the set of time-lags corresponding to the autocorrelation matrices was set to $\mathcal{M}=\{1,2,\ldots,30\}$ across all compared methods.

In the MT-SOBI algorithm, the functions $h(\cdot)$ and $g(\cdot)$ shaping the parametric MT-functions $u(\cdot,\cdot;\theta)$ (17) and $v(\cdot,\cdot;\theta)$ (20) were set to $h(r)=\exp(-r^2)$ and $g(\mathbf{r})=\exp(-\|\mathbf{r}\|^2)$, respectively. As discussed below (17) and (20), these choices yield strongly outlier-suppressive MT-functions that satisfy conditions (6) and (7). The scaling parameter θ was chosen according to (18). The approximate joint diagonalization (AJD) of the symmetrized empirical MT-autocorrelation matrices (19) was performed using the ACDC method [16].

The MF-SOBI [10] algorithm was implemented with a 10-sample-wide median filtering applied to each component sequence $\{x_{i,n}\}_{n=1}^N$, for $i=1,\ldots,p$, separately.

C. Results

Figs. 1 and 2 depict the averaged Amari error as a function of GSNR for the Gaussian and t-distributed noise, respectively. Notably, MF-SOBI exhibits the weakest performance due to the non-linear median filtering, which distorts the linear mixing structure (8). In contrast, MT-SOBI achieves the best performance in both noise scenarios.

In the Gaussian noise case, the performance advantage of MT-SOBI over SOBI, SAM-SOBI, and eSAM-SOBI may be attributed to the fact that, unlike these methods, which apply pre-whitening and propagate estimation errors through the entire separation process, MT-SOBI avoids this step, thereby preserving the original data structure. Although estimation errors in the projection matrix may still propagate, they remain localized within the scalar data-weighting MT-function rather than affecting the full data transformation. Furthermore, the performance advantage over SAM-SOBI and eSAM-SOBI may be attributed to the property that the MT-autocorrelation matrices preserve the spatial correlation structure under the additive noise model, as evidenced by (15). This property is necessary to ensure consistent estimation of $\mathbf{A}$ through AJD.

For the heavy-tailed t-distributed noise, the performance advantage of MT-SOBI over the non-robust SOBI is anticipated. The advantage over SAM-SOBI and eSAM-SOBI may stem from three factors: (i) avoidance of pre-whitening, (ii) spatial correlation structure preservation in the presence of additive noise, and (iii) more effective outlier suppression. The latter arises from the fact that unlike the empirical sign-autocorrelation matrices used in SAM-SOBI and eSAM-SOBI, the empirical MT-autocorrelation matrices reject large-norm outliers under the considered MT-functions.

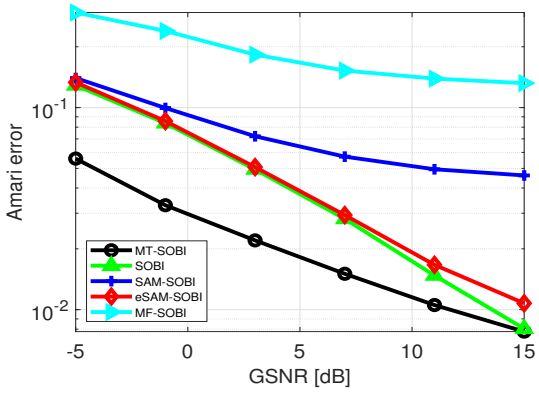

Fig. 1. Gaussian noise: Averaged Amari error versus GSNR.

VII. Conclusion

In this paper, we introduced a new robust variant of the SOBI algorithm. The proposed method, called MT-SOBI, operates by transforming pairwise joint probability measures of time-lagged samples. We showed that this framework yields superior performance compared to other robust SOBI extensions. Future research will explore refined scaling of the MT-functions to optimize performance-related objectives. Additionally, a recursive version of this method will be investigated.

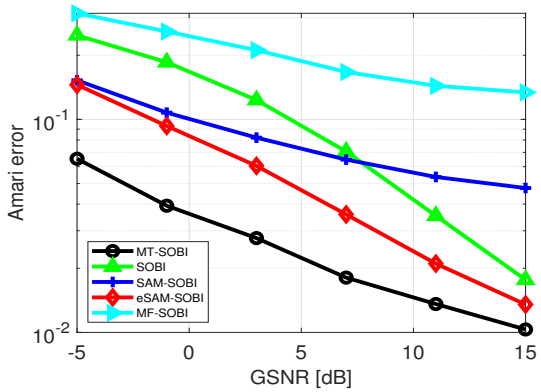

Fig. 2. t-distributed noise: Averaged Amari error versus GSNR.

References

[1] P. Common and C. Jutten, *Handbook of blind source separation: independent component analysis and applications,* Academic press, 2010.

[2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434-444, 1997.

[3] F. J. Theis, N. S. Müller, C. Plant, C. Böhm, V. Koivunen, and H. Oja, "Robust second-order source separation identifies experimental responses in biomedical imaging," in *Proc. LVA/ICA*, pp. 466-473, 2010.

[4] P. Ilmonen, K. Nordhausen, H. Oja, and F. Theis, "An affine equivariant robust second-order BSS method," in *Proc. LVA/ICA*, pp. 328-335, 2015.

[5] S. Visuri, V. Koivunen, and H. Oja, "Sign and rank covariance matrices," *Journal of Statistical Planning and Inference*, vol. 91, no. 2, pp. 557-575, 2000.

[6] T. P. Hettmansperger and R. H. Randles, "A practical affine equivariant multivariate median," *Biometrika*, vol. 89, no. 4, pp. 851-860, 2002.

[7] P. J. Huber, *Robust Statistics*: Wiley, 1981.

[8] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*: Wiley, 2011.

[9] K. Todros and A. O. Hero, "Robust multiple signal classification via probability measure transformation," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1156-1170, 2015.

[10] F. Miao, R. Zhao, L. Jia, and X. Wang, "Fault diagnosis of rotating machinery based on multi-sensor signals and median filter second-order blind identification (MF-SOBI)," *Applied Sciences*, vol. 10, 2020.

[11] Z. Bekhtaoui, K. Abed-Meraim, A. Meche, and M. Thameri, "A new robust adaptive algorithm for second-order blind source separation," *ENP Engineering Science Journal*, vol. 2, no. 1, pp. 21-28, 2022.

[12] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results, and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597-5625, 2012.

[13] K. Todros and A. O. Hero, "Measure-transformed quasi-maximum likelihood estimation," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 748-763, 2016.

[14] T. Ben-Guy and K. Todros, "Blind separation of noisy piecewise-stationary mixtures via probability measure transform," *Signal Processing*, vol. 208, p. 108967, 2023.

[15] N. Halay and K. Todros, "MSE-based optimization of the measure-transformed MUSIC algorithm," *Signal Processing*, vol. 160, pp. 150-163, 2019.

[16] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1545-1553, 2002.

[17] G. B. Folland, *Real Analysis*: Wiley, 1984.

[18] R. A. Horn and C. R. Johnson, *Matrix analysis.* Cambridge Univ. Press, 2004.

[19] S. I. Amari, A. Cichocki and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 8, 1995.