

Exact Algebraic Blind Source Separation Using Side Information

Amir Weiss

Dept. of Computer Science and Applied Mathematics
Weizmann Institute of Science
amir.weiss@weizmann.ac.il

Arie Yeredor

School of Electrical Engineering
Tel-Aviv University
arie@eng.tau.ac.il

Abstract—Classical Blind Source Separation (BSS) methods rarely attain exact separation, even under noiseless conditions. In addition, they often rely on particular structural or statistical assumptions (e.g., mutual independence) regarding the sources. In this work we consider a (realistic) “twist” in the classical linear BSS plot, which, quite surprisingly, not only enables perfect separation (under noiseless conditions), but is also free of any assumptions (except for regularity assumptions) regarding the sources or the mixing matrix. In particular, we consider the standard linear mixture model, augmented by a single ancillary, unknown linear mixture of some known linear transformations of the sources. We derive a closed-form expression for an exact algebraic solution, free of any statistical considerations whatsoever, attaining perfect separation in the noiseless case. In addition, we propose a well-behaved solution for the same model in the presence of noise or other measurement inaccuracies. Our derivations are corroborated by several simulation results.

Index Terms—Blind source separation, algebraic methods, side information, total least squares.

I. INTRODUCTION

Consider K different sources (e.g., transmitters), spatially spread out and positioned at known locations, transmitting different signals. Now, assume that these signals are received at a single site by K different sensors (e.g., antennas) with different (unknown, or uncalibrated) radiation patterns, directed at different directions, not necessarily in the form of a calibrated phased array. The signal received at each antenna is then a different, unknown linear mixture of all K sources.

Now assume a second, remote reception site with a *single* (possibly different) sensor. The signal received at this site is an unknown mixture of differently-delayed versions of the same sources, due to propagation delays. Knowing the positions of the sources and of the sensor sites, these delays can be calculated, and in turn define a set of K deterministic, known, linear transformations capturing the linear relations between the received signals at the primary site and the (time-shifted) received signals at the remote site. Thus, the $K + 1$ mixtures from both sites give rise to a Blind Source Separation (BSS) problem with side information, in the form of K linear transformations linking the different signals from the same sources at these two spatially separated sites.

The scenario described above, illustrated in Fig. 1, is merely a particular case of the more general model for the BSS problem we address in this paper. In fact, apart from the side information assumed to be known *a-priori*, our model

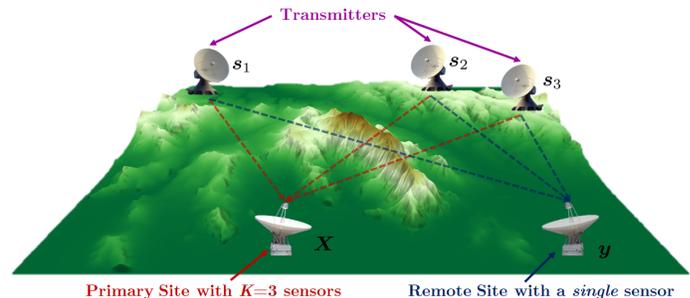


Fig. 1: Illustration of the scenario described in the beginning of Section I. This is (one) particular case of the model for the more general BSS problem we address in this paper. Here, $\{s_i\}$ are the sources, \mathbf{X} is the matrix of observed mixtures from the first (primary) site, and \mathbf{y} is the single (vector) observed mixture from the remote site.

is quite general and robust relative to most classical BSS approaches [1]. In particular, the general appeal of our model builds upon the fact that we do not assume (almost) anything regarding the underlying latent sources. These signals may (or may not) be modeled as stochastic—for example, non-Gaussian or Gaussian, all having the same (known or unknown) spectrum, etc.—or as deterministic, real- or complex-valued. This is unlike many other BSS approaches, in which (at least) some *a-priori* assumptions are made in this respect. For example, in the Independent Component Analysis (ICA, [2]) framework, the sources are modeled as mutually statistically independent stochastic processes. Within this framework, an abundance of algorithms for various signal models have been developed, such as the Joint Approximation Diagonalization of Eigen-matrices (JADE) and FastICA algorithms for non-Gaussian signals ([3] and [4], resp.), the Second-Order Blind Identification (SOBI) and weights-adjusted SOBI algorithms for stationary temporally-diverse sources ([5] and [6], resp.) and the maximum likelihood based minimum mean square error separation of stationary Gaussian sources [7], to name but a few for different signal models. Different approaches for BSS have also been considered, such as independent subspace analysis [8] and BSS using sparse representations [9]. Nonetheless, as already pointed out, such methods are developed based on some presumed (for the most part, statistical) properties of the unobserved sources. In contrast, for the model under consideration here, only the “sites-linking” linear transformations are required to be known *a-priori*. Note that a successful separation algorithm under this “non-assumption”

regarding the “nature” of the sources is quite a powerful tool.

Our main contribution in this work is an exact algebraic solution providing *perfect* separation (in the noiseless case), namely we derive a closed-form expression for a unique, scaled, permutation-free unmixing matrix. Consequently, exact reconstruction of the sources is enabled, up to some unique (inevitable) scaling. We stress that here, an “algebraic solution” should be interpreted as being free of any statistical considerations whatsoever, as opposed to, e.g., methods presented in [1], Chapter 5, operating on (estimated) higher-order cumulant tensors. In addition, we also provide a closed-form expression for a noisy case solution, which may be invoked, in lieu of the noiseless (exact) solution, in low Signal-to-Noise Ratio (SNR).

A. Notations

We use a , \mathbf{a} and \mathbf{A} for a scalar, column vector and matrix, resp., where A_{ij} denotes the (i, j) -th element of the matrix \mathbf{A} . Likewise, we denote the all-zeros vector and matrix as $\mathbf{0}$ and \mathbf{O} (with context-implied dimensions), resp. The superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote the transposition, Hermitian transposition and inverse operators, resp., and $\|\cdot\|_F$ denotes the Frobenius norm. We denote by \mathbf{I}_K the $K \times K$ identity matrix, and the pinning vector \mathbf{e}_k denotes the k -th column of \mathbf{I}_K . The Kronecker and Khatri-Rao (column-wise Kronecker) products are denoted by \otimes and \diamond , resp. We also define $\text{vec}(\cdot)$ as the operator which concatenates the columns of an $M \times N$ matrix into an $MN \times 1$ column vector. Lastly, we define the operator $\text{Diag}(\cdot)$, which creates an $N \times N$ diagonal matrix from its N -dimensional vector argument.

II. PROBLEM FORMULATION

Consider the following static, instantaneous, linear model

$$1 \leq k \leq K: \mathbf{x}_k = \sum_{\ell=1}^K A_{\ell k} \mathbf{s}_\ell \in \mathbb{C}^{N \times 1} \Rightarrow \mathbf{X} = \mathbf{S}\mathbf{A} \in \mathbb{C}^{N \times K}, \quad (1)$$

where $\mathbf{S} \triangleq [\mathbf{s}_1 \cdots \mathbf{s}_K] \in \mathbb{C}^{N \times K}$ denotes a matrix of K unobservable source signals of length N , $\mathbf{A} \in \mathbb{C}^{K \times K}$ is an unknown mixing matrix, and the observed mixture signals are given by $\mathbf{X} \triangleq [\mathbf{x}_1 \cdots \mathbf{x}_K] \in \mathbb{C}^{N \times K}$ (for convenience of the exposition, we prefer this column-wise mixture notation over the classical row-wise mixture notation $\mathbf{X} = \mathbf{A}\mathbf{S}$ with transposed definitions of \mathbf{S} , \mathbf{A} and \mathbf{X}). Our sole assumption regarding the model equation (1) is that the mixing matrix \mathbf{A} is invertible (a very common assumption in BSS). Next, consider yet another *single* ancillary mixture given by

$$\mathbf{y} = \sum_{\ell=1}^K c_\ell \mathbf{T}_\ell \mathbf{s}_\ell \in \mathbb{C}^{N \times 1}, \quad (2)$$

where $\mathbf{c} \triangleq [c_1 \cdots c_K]^T \in \mathbb{C}^{K \times 1}$ is a vector of unknown, non-zero mixing coefficients, and $\{\mathbf{T}_k \in \mathbb{C}^{N \times N}\}_{k=1}^K$ are known linear transformations. Notice that \mathbf{y} is a different mixture of some (general) linearly-transformed versions of the same sources comprising the mixtures $\mathbf{x}_1, \dots, \mathbf{x}_K \in \mathbb{C}^{N \times 1}$ (i.e., the columns of \mathbf{X}). Referring to the example presented in Section I (Fig. 1), in that scenario, the matrix \mathbf{T}_k would represent

the known (relative) time-shift between the signal received from the k -th source at the primary site to the signal from the same source received at the remote site. Notice that a time-shift is, in particular, an instance of a Linear Time-Invariant (LTI) system parametrized by a single parameter. Therefore, knowing these K time-shifts (l parameters) is equivalent to knowing these K linear transformations¹. Moreover, since the \mathbf{T}_k matrices are not confined to any particular structure, given that the number of observations N is sufficiently large, they can represent any Finite Impulse Response (FIR) LTI system, or, furthermore, any linear (possibly time-varying) system. In our BSS context, the ancillary observed mixture (2) together with the knowledge of the K linear transformations $\{\mathbf{T}_k\}_{k=1}^K$ are referred to as “side information”.

Thus, the problem at hand is as follows:

Problem: Given \mathbf{X} , \mathbf{y} and $\{\mathbf{T}_k\}_{k=1}^K$, find an unmixing matrix $\bar{\mathbf{B}}$, such that $\mathbf{A}\bar{\mathbf{B}} = \text{Diag}(\boldsymbol{\gamma})$, for some $\boldsymbol{\gamma} \in \mathbb{C}^{K \times 1}$.

In other words, our goal is to find some scaled version of \mathbf{A}^{-1} , i.e., $\bar{\mathbf{B}} = \mathbf{A}^{-1} \text{Diag}(\boldsymbol{\gamma})$, with which the sources may be perfectly separated—by multiplying \mathbf{X} by $\bar{\mathbf{B}}$ on the right—up to a scaling coefficient $\gamma_k \neq 0$ of the k -th source.

III. THE EXACT ALGEBRAIC SEPARATING SOLUTION

We begin with a short description of our strategy for the solution. The key aspect is to first take notice that, according to (1) and (2), \mathbf{X} and \mathbf{y} are linear (even if unknown) transformations of the same source signals $\mathbf{s}_1, \dots, \mathbf{s}_K$. Therefore, based on the known transformations $\{\mathbf{T}_k\}_{k=1}^K$, we shall work towards formulating the linear relation between \mathbf{y} and \mathbf{X} in terms of the desired true unmixing matrix $\mathbf{B} \triangleq \mathbf{A}^{-1}$.

Let us start by defining the ℓ -transformed k -th source

$$\tilde{\mathbf{s}}_k^{(\ell)} \triangleq \mathbf{T}_\ell \mathbf{s}_k \in \mathbb{C}^{N \times 1}, \quad \forall k, \ell \in \{1, \dots, K\}, \quad (3)$$

where we have K^2 such transformed sources. Accordingly, let

$$\tilde{\mathbf{S}}^{(k)} \triangleq \mathbf{T}_k \mathbf{S} = [\tilde{\mathbf{s}}_1^{(k)} \cdots \tilde{\mathbf{s}}_K^{(k)}] \in \mathbb{C}^{N \times K}, \quad \forall k \in \{1, \dots, K\}, \quad (4)$$

and further define the (realizable) transformed mixtures

$$\tilde{\mathbf{X}}^{(k)} \triangleq \mathbf{T}_k \mathbf{X} \in \mathbb{C}^{N \times K}, \quad \forall k \in \{1, \dots, K\}. \quad (5)$$

Collecting all $\{\tilde{\mathbf{S}}^{(k)}\}$ and $\{\tilde{\mathbf{X}}^{(k)}\}$, we define the augmented transformed sources and mixtures matrices, resp.,

$$\tilde{\mathbf{S}} \triangleq [\tilde{\mathbf{S}}^{(1)} \cdots \tilde{\mathbf{S}}^{(K)}] \in \mathbb{C}^{N \times K^2}, \quad (6)$$

$$\tilde{\mathbf{X}} \triangleq [\tilde{\mathbf{X}}^{(1)} \cdots \tilde{\mathbf{X}}^{(K)}] \in \mathbb{C}^{N \times K^2}. \quad (7)$$

¹Assuming the relative delays are an integer multiple of the sampling period. Otherwise, if the signals are bandlimited, sampled at least at their Nyquist rate, and N is sufficiently large, \mathbf{T}_k can be a matrix-form FIR approximation of the (accurate) equivalent discrete-time interpolator realizing the delay up to negligible edge effects [10].

Clearly, (7) is (still) a mixture of (6) with

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{S}} \begin{bmatrix} \mathbf{A} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{A} & \dots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \dots & \mathbf{A} \end{bmatrix} = \widetilde{\mathbf{S}} \underbrace{(\mathbf{I}_K \otimes \mathbf{A})}_{\triangleq \widetilde{\mathbf{A}} \in \mathbb{C}^{K^2 \times K^2}} = \widetilde{\mathbf{S}} \widetilde{\mathbf{A}}. \quad (8)$$

Moving to the second model equation (2), notice that the observed ancillary mixture reads

$$\begin{aligned} \mathbf{y} &= \sum_{\ell=1}^K c_\ell \underbrace{\mathbf{T}_\ell \mathbf{s}_\ell}_{\widetilde{\mathbf{s}}_\ell^{(\ell)}} = [\widetilde{\mathbf{s}}_1^{(1)} \dots \widetilde{\mathbf{s}}_K^{(K)}] \mathbf{c} \\ &= \widetilde{\mathbf{S}} \begin{bmatrix} e_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & e_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & e_K \end{bmatrix} \mathbf{c} = \widetilde{\mathbf{S}} \underbrace{(\mathbf{I}_K \diamond \mathbf{I}_K)}_{\triangleq \widetilde{\mathbf{J}}_K \in \mathbb{R}^{K^2 \times K}} \mathbf{c} = \widetilde{\mathbf{S}} \widetilde{\mathbf{J}}_K \mathbf{c}. \end{aligned} \quad (9)$$

Summarizing the relations we have obtained so far, we have

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{S}} \widetilde{\mathbf{A}}, \quad (10)$$

$$\mathbf{y} = \widetilde{\mathbf{S}} \widetilde{\mathbf{J}}_K \mathbf{c}. \quad (11)$$

Since \mathbf{A} is invertible by assumption, so is $\widetilde{\mathbf{A}}$. Thus, denote

$$\widetilde{\mathbf{B}} \triangleq \widetilde{\mathbf{A}}^{-1} = (\mathbf{I}_K \otimes \mathbf{B}), \quad (12)$$

to write (10) in terms of $\widetilde{\mathbf{B}}$, leading to

$$\widetilde{\mathbf{X}} \widetilde{\mathbf{B}} = \widetilde{\mathbf{S}} \implies \mathbf{y} = \widetilde{\mathbf{X}} \widetilde{\mathbf{B}} \widetilde{\mathbf{J}}_K \mathbf{c}. \quad (13)$$

Yet, observe that

$$\widetilde{\mathbf{B}} \widetilde{\mathbf{J}}_K \mathbf{c} = (\mathbf{I}_K \otimes \mathbf{B}) (\mathbf{I}_K \diamond \mathbf{I}_K) \mathbf{c} = (\mathbf{I}_K \diamond \mathbf{B}) \mathbf{c} \quad (14)$$

$$= \text{vec}(\mathbf{B} \text{Diag}(\mathbf{c})) \triangleq \mathbf{b}_c \in \mathbb{C}^{K^2 \times 1}, \quad (15)$$

where we have used the properties (e.g., [11], Section II)

$$(\mathbf{Z} \otimes \mathbf{W}) (\mathbf{Q} \diamond \mathbf{P}) = (\mathbf{Z}\mathbf{Q}) \diamond (\mathbf{W}\mathbf{P}), \quad [\text{mixed product rule}]$$

$$(\mathbf{Z} \diamond \mathbf{W}) \mathbf{q} = \text{vec}(\mathbf{W} \text{Diag}(\mathbf{q}) \mathbf{Z}).$$

Therefore, assuming $\widetilde{\mathbf{X}}$ is full column rank, substituting \mathbf{b}_c , which is simply the vectorization of a scaled version of \mathbf{B} (i.e., our desired unknown), into (13) readily gives

$$\mathbf{y} = \widetilde{\mathbf{X}} \mathbf{b}_c \implies \mathbf{b}_c = \left(\widetilde{\mathbf{X}}^H \widetilde{\mathbf{X}} \right)^{-1} \widetilde{\mathbf{X}}^H \mathbf{y}. \quad (16)$$

Rearranging \mathbf{b}_c , given by (16), in a $K \times K$ matrix gives us $\mathbf{B} \text{Diag}(\mathbf{c})$, with which $\mathbf{A} \mathbf{B} \text{Diag}(\mathbf{c}) = \text{Diag}(\mathbf{c})$, as desired.

Remarks:

- i. Notice that our solution provides exact separation of the sources, i.e., with zero Interference-to-Source Ratio (ISR, for the definition see (27) in Section V).
- ii. The condition that $\widetilde{\mathbf{X}}$ is full column rank (namely, its rank is K^2) is equivalent to a requirement that $\widetilde{\mathbf{S}}$ is full column rank, since they are related via (10) (which is invertible by virtue of the assumed invertibility of \mathbf{A}). This condition may be regarded as the “identifiability condition” for this side-information-aided scenario.

iii. Note yet another interesting observation: as long as N is sufficiently larger than K^2 , invertibility of any of the transformations \mathbf{T}_k is *not* required. Indeed, nowhere in our derivations have we used the inverse of \mathbf{T}_k .

It is important to keep in mind, that although this perfectly separating solution assumes a “clean” model (free of noise or of other measurement errors), in classical BSS perfect separation is usually unattainable, not even with a “clean” model. However, in order to address more realistic scenarios, we now turn to an extended version of our model, incorporating possible (additive) errors in equations (1) and (2).

IV. MITIGATING MEASUREMENT ERRORS

We begin by introducing measurement errors $\boldsymbol{\varepsilon}_y \in \mathbb{C}^{N \times 1}$ to the ancillary mixture (2) only, i.e.,

$$\mathbf{y}_\varepsilon \triangleq \mathbf{y} + \boldsymbol{\varepsilon}_y = \widetilde{\mathbf{X}} \mathbf{b}_c + \boldsymbol{\varepsilon}_y. \quad (17)$$

It is readily seen that, as long as (10) holds, the exact algebraic solution (16) coincides with the Least Squares (LS) estimate (with \mathbf{y}_ε replacing \mathbf{y}), namely

$$\left(\widehat{\mathbf{b}}_c \right)_{\text{LS}} = \left(\widetilde{\mathbf{X}}^H \widetilde{\mathbf{X}} \right)^{-1} \widetilde{\mathbf{X}}^H \mathbf{y}_\varepsilon. \quad (18)$$

Of course, if $\boldsymbol{\varepsilon}_y$ is a random vector with a known mean vector and covariance matrix, the optimally-weighted, unbiased LS estimate may be easily calculated as well (e.g., [12]).

Having shown that (16) is also a remedy for inaccuracies in (2), let us consider the case where (1) is inaccurate as well. Formally, we consider the extended model

$$\mathbf{X}_\varepsilon \triangleq \mathbf{X} + \boldsymbol{\varepsilon}_x = \mathbf{S} \mathbf{A} + \boldsymbol{\varepsilon}_x, \quad (19)$$

$$\mathbf{y}_\varepsilon = \sum_{\ell=1}^K c_\ell \mathbf{T}_\ell \mathbf{s}_\ell + \boldsymbol{\varepsilon}_y = \widetilde{\mathbf{S}} \widetilde{\mathbf{J}}_K \mathbf{c} + \boldsymbol{\varepsilon}_y. \quad (20)$$

One can easily repeat the same derivation as in Section III to obtain the transformed model equation

$$\widetilde{\mathbf{X}}_\varepsilon \triangleq \widetilde{\mathbf{S}} \widetilde{\mathbf{A}} + \widetilde{\boldsymbol{\varepsilon}}_x \implies \widetilde{\mathbf{S}} = \left(\widetilde{\mathbf{X}}_\varepsilon - \widetilde{\boldsymbol{\varepsilon}}_x \right) \widetilde{\mathbf{B}}, \quad (21)$$

with

$$\forall k \in \{1, \dots, K\}: \widetilde{\boldsymbol{\varepsilon}}_x^{(k)} \triangleq \mathbf{T}_k \boldsymbol{\varepsilon}_x \implies \widetilde{\boldsymbol{\varepsilon}}_x \triangleq \left[\widetilde{\boldsymbol{\varepsilon}}_x^{(1)} \dots \widetilde{\boldsymbol{\varepsilon}}_x^{(K)} \right]. \quad (22)$$

Substituting (21) into (20) yields

$$(\mathbf{y}_\varepsilon - \boldsymbol{\varepsilon}_y) = \left(\widetilde{\mathbf{X}}_\varepsilon - \widetilde{\boldsymbol{\varepsilon}}_x \right) \mathbf{b}_c, \quad (23)$$

which gives rise to a TLS problem [13]. Thus, if we denote

$$\left[\widetilde{\mathbf{X}}_\varepsilon \quad \mathbf{y}_\varepsilon \right] \triangleq \left[\mathbf{U}_b \quad \mathbf{u}_\varepsilon \right] \begin{bmatrix} \boldsymbol{\Sigma}_b & \mathbf{0} \\ \mathbf{0}^T & \sigma_\varepsilon \end{bmatrix} \begin{bmatrix} \mathbf{V}_{bb} & \mathbf{v}_{b\varepsilon} \\ \mathbf{v}_{\varepsilon b}^T & v_\varepsilon \end{bmatrix} \in \mathbb{C}^{N \times (K^2+1)} \quad (24)$$

as the Singular Value Decomposition (SVD) of $\left[\widetilde{\mathbf{X}}_\varepsilon \quad \mathbf{y}_\varepsilon \right]$, then

$$\left(\widehat{\mathbf{b}}_c \right)_{\text{TLS}} = -\frac{\mathbf{v}_{b\varepsilon}}{v_\varepsilon} \quad (25)$$

is the well-known TLS estimate, solving

$$\left(\widehat{\mathbf{b}}_c \right)_{\text{TLS}} = \underset{\mathbf{b}_c \in \mathbb{C}^{K^2 \times 1}}{\text{argmin}} \left\| \left[\widetilde{\boldsymbol{\varepsilon}}_x \quad \boldsymbol{\varepsilon}_y \right] \right\|_{\text{F}}^2 \quad (26)$$

$$\text{subject to: } (\mathbf{y}_\varepsilon - \boldsymbol{\varepsilon}_y) = \left(\widetilde{\mathbf{X}}_\varepsilon - \widetilde{\boldsymbol{\varepsilon}}_x \right) \mathbf{b}_c.$$

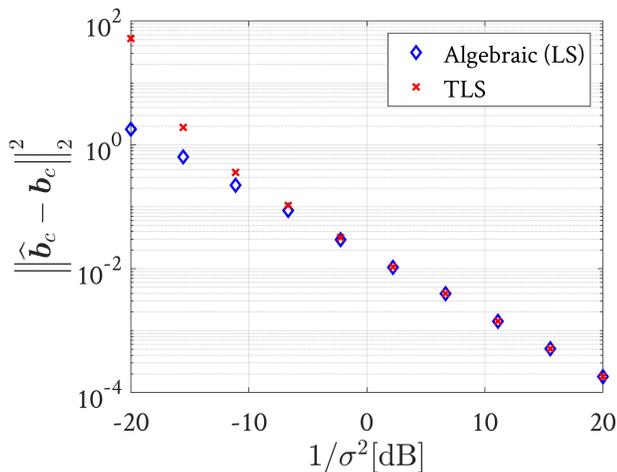


Fig. 2: Experiment 1: Average ℓ^2 -norm of $(\hat{\mathbf{b}}_c - \mathbf{b}_c)$ vs. $1/\sigma^2$, for $K = 2$, $N = 10$. Results were obtained by averaging 10^3 independent trials.

Lastly, notice that all the results presented in this section are also valid and hold true for the case of more sensors than sources at the primary site, where the mixing matrix is not a square matrix, i.e., $\mathbf{A} \in \mathbb{C}^{K \times M}$ with $M > K$ (as long as it is full column rank).

V. SIMULATION RESULTS

It is easy to verify by a simple numerical computation that the exact algebraic solution (16) is indeed correct. Several Matlab[®] computations, in which \mathbf{S} , \mathbf{A} , $\{\mathbf{T}_k\}_{k=1}^K$ and \mathbf{c} were drawn independently from the standard complex normal distribution (for several, different values of K and N), resulted in an “error vector” $(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_c)$ with a squared norm (normalized by N) of the order of $\sim 10^{-28}$, which is clearly due to machine accuracy limitations, and therefore implies perfect separation in the noiseless case (since \mathbf{b}_c is a vector-form of an *exact* separating matrix). Therefore, in this section, we shall focus on empirical performance analysis of the proposed solutions in the presence of measurement errors.

In our first simulation experiment, we introduce measurement errors only to the ancillary mixture \mathbf{y} . We consider a mixture of $K = 2$ sources of length $N = 10$. The mixing coefficients, i.e., the elements of \mathbf{A} and \mathbf{c} , the K linear transformations $\{\mathbf{T}_k\}_{k=1}^K$ and the source signals \mathbf{S} were all drawn (once, and then fixed) independently from the standard complex normal distribution. In each independent trial out of such 10^3 , the elements of ε_y were drawn independently from the zero-mean complex normal distribution with variance σ^2 . Notice that in this experiment, we wish to examine the performance of the proposed estimates in a scenario where, except for ε_y , all the parameters are considered deterministic, and therefore, are drawn once and then fixed. Fig. 2 shows the average empirical (squared) ℓ^2 -norm of the error vector $(\hat{\mathbf{b}}_c - \mathbf{b}_c)$ of the algebraic (LS) and the TLS estimates vs. σ^2 . It is evident from the constant rate of decay that both solutions yield perfect separation as $\sigma^2 \rightarrow 0$. Furthermore, as expected, the algebraic solution is superior in the low SNR regime, and the TLS solution converges to the former as the SNR increases.

In our second experiment, we simulate the more practical scenario illustrated in Fig. 1 with $K = 3$ sources emitting communication signals. In particular, we consider Orthogonal Frequency Division Multiplexing (OFDM, e.g., [14]) signals, from a unit power 8-Phase Shift Keying (8-PSK) constellation. Note that due to the (near) Gaussianity and spectral similarity of such OFDM signals, the mixture at the primary site is (almost) unidentifiable (and thus practically inseparable) using classical ICA methods, even with independent sources. And yet, thanks to the availability of just a single mixture signal at the remote site, and knowledge of the associated delays, we are able to separate the otherwise non-separable sources. In each trial, N 8-PSK symbols were drawn independently and uniformly for every source. Thus, each column of \mathbf{S} , the mixed source signals at the primary site, is the (normalized) length- N inverse FFT of the vector of symbols. The mixed source signals received at the remote site are differently-delayed versions of those from the primary site, with known (integer) delays $\boldsymbol{\tau} \triangleq [2 \ 5 \ 11]^T$ [samples], where τ_k is the delay associated with the k -th source signal. Thus, these known delays imply the known K linear transformations

$$(\mathbf{T}_k)_{n_1 n_2} = \begin{cases} 1, & n_1 - n_2 = \tau_k \\ 0, & n_1 - n_2 \neq \tau_k \end{cases}, \quad \forall n_1, n_2 \in \{1, \dots, N\}, \quad \forall k \in \{1, 2, 3\}.$$

The mixing matrix \mathbf{A} was generated as in the previous experiment, and the mixing coefficients \mathbf{c} were generated according to $c_k = e^{-\lambda_k \tau_k}$, where $\lambda_k \sim U(0.9, 1.1)$ are some unknown decay rates (drawn once, and then fixed). In addition, the measurements are contaminated by additive temporally- and spatially-white complex normal noise, with variances σ_x^2 and σ_y^2 for the elements of \mathcal{E}_x and ε_y , resp.

Note that in this scenario, for any PSK constellation, a scaling ambiguity by a positive (real-valued) scalar in the separated, estimated sources is immaterial w.r.t. the Minimum Euclidean Distance (MED) estimation rule, and in turn, in terms of the resulting Symbol Error Rate (SER). Therefore, in each trial, we estimate the separating matrix $\mathbf{B} \text{Diag}(\mathbf{c})$ and apply zero-forcing ([15]) equalization to obtain estimates of the sources, from which the transmitted symbols are estimated via FFT followed by a decision according to the MED rule.

Fig. 3 shows the empirical SER manifold vs. the noise variances obtained with each of the proposed separating solutions, with $N = 2^4$ fixed. It is clearly seen that the algebraic estimate (which is no longer the LS estimate in the presence of \mathcal{E}_x) dominates the TLS solution for a relatively “small” value of N . Moreover, the simple algebraic solution may be obtained with lower computational complexity ($K^2 \times K^2$ matrix inversion vs. $N \times (K + 1)$ matrix SVD). Still, both solutions yield decaying SER curves with $\sigma_x^2, \sigma_y^2 \rightarrow 0$ for a fixed (even if low) N .

We continue to evaluate the performance in term of the ISR,

$$\text{ISR}_{k\ell} \triangleq \mathbb{E} \left[\frac{|(\widehat{\mathbf{A}\mathbf{B}})_{k\ell}|^2}{|(\widehat{\mathbf{A}\mathbf{B}})_{kk}|^2} \right], \quad \forall k \neq \ell \in \{1, \dots, K\}, \quad (27)$$

quantifying the quality of separation of an unmixing matrix estimate $\widehat{\mathbf{B}}$ (where $\mathbb{E}[\cdot]$ denotes expectation). Fig. 4 presents

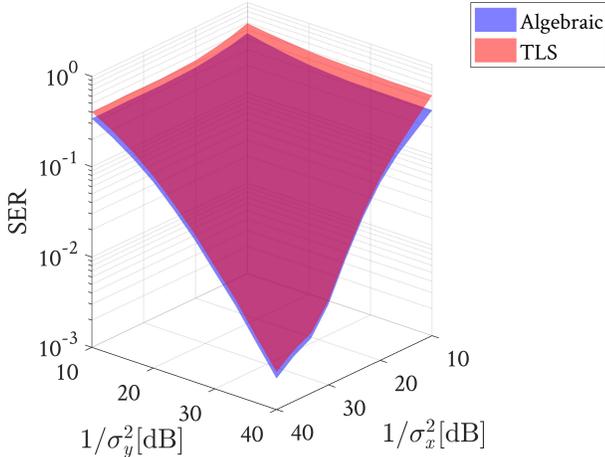


Fig. 3: Experiment 2: Empirical SER manifold vs. σ_x^2 and σ_y^2 for $N = 2^4$ with 8-PSK OFDM signals. Results are based on 10^4 independent trials.

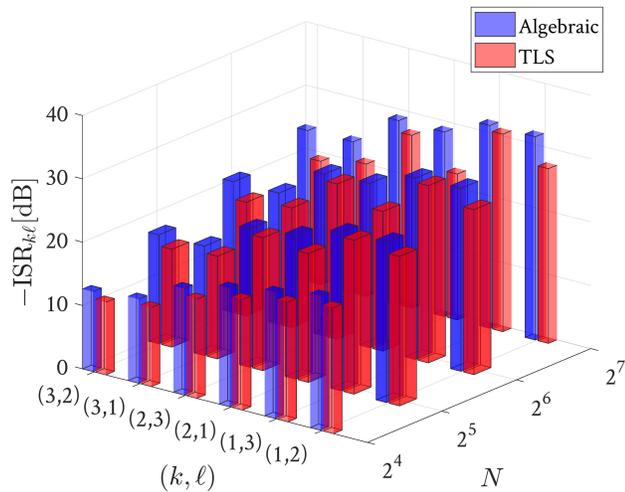


Fig. 4: Experiment 2: *minus* Average ISR [the higher the better] vs. the FFT size, N , for $\sigma_x^2 = -20$ [dB] and $\sigma_y^2 = -10$ [dB]. The algebraic (exact) solution for the error-free model provides better separation when σ_x^2 is sufficiently small. Results were obtained by averaging 10^4 independent trials.

the (minus) average ISR (in [dB]) of the proposed estimates vs. the FFT size N for fixed $\sigma_y^2 = -10$ [dB] and $\sigma_x^2 = -20$ [dB]. Evidently, when the SNR is sufficiently high at the primary site (i.e., $\sigma_x^2 \ll 1$), the TLS estimate is inferior to the algebraic solution, which performs better for noisier side information \mathbf{y}_ε . Of course, when the mixture \mathbf{X}_ε is too noisy, the TLS solution is preferable, as demonstrated in Fig. 5, presenting the average empirical squared ℓ^2 -norm of $(\hat{\mathbf{b}}_c - \mathbf{b}_c)$ (normalized by N) vs. $1/\sigma^2$, with $N = 2^8$ and $\sigma^2 = \sigma_y^2 = \sigma_x^2 - 10$ [dB].

VI. CONCLUSION

Within the scope of BSS, we presented a model with side-information in the form of a single ancillary mixture and linear transformations which link the sources in both mixtures. For this model, we derived an algebraic closed-form solution, providing perfect separation in the noiseless case. In addition, we derived the TLS solution, which serves as a remedy for (possible) additive errors. Empirical simulation results reflect

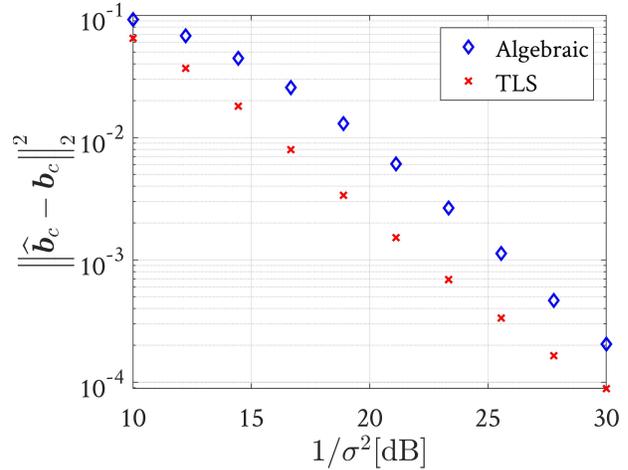


Fig. 5: Experiment 2: Average ℓ^2 -norm of $(\hat{\mathbf{b}}_c - \mathbf{b}_c)$ (normalized by N) vs. $1/\sigma^2$, for $K = 3$, $N = 2^8$ and $\sigma^2 = \sigma_y^2 = \sigma_x^2 - 10$ [dB]. The TLS may be superior to the algebraic solution under some conditions. Results were obtained by averaging 10^3 independent trials.

the effectiveness of both solutions, each under its respective appropriate conditions.

REFERENCES

- [1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic press, 2010.
- [2] P. Comon, "Independent component analysis, a new concept?" *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [3] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," in *IEE proceedings F (radar and signal processing)*, vol. 140, no. 6. IET, 1993, pp. 362–370.
- [4] A. Hyvarinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
- [5] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A Blind Source Separation Technique Using Second-Order Statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.
- [6] A. Yeredor, "Blind Separation of Gaussian Sources via Second-Order Statistics with Asymptotically Optimal Weighting," *IEEE Signal Processing Letters*, vol. 7, no. 7, pp. 197–200, 2000.
- [7] A. Weiss and A. Yeredor, "A Maximum Likelihood-Based Minimum Mean Square Error Separation and Estimation of Stationary Gaussian Sources From Noisy Mixtures," *IEEE Trans. on Signal Processing*, vol. 67, no. 19, pp. 5032–5045, 2019.
- [8] M. A. Casey and A. Westner, "Separation of Mixed Audio Sources By Independent Subspace Analysis," in *ICMC*, 2000, pp. 154–161.
- [9] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal processing*, vol. 81, no. 11, pp. 2353–2362, 2001.
- [10] A. Yeredor, "Blind Source Separation with Pure Delay Mixtures," in *Proc. ICA*, 2001, pp. 522–527.
- [11] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor Decomposition for Signal Processing and Machine Learning," *IEEE Trans. on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.
- [12] A. Weiss and A. Yeredor, "Asymptotically Optimal Recovery of Gaussian Sources from Noisy Stationary Mixtures: The Least-Noise Maximally-Separating Solution," in *Proc. of ICASSP*, 2019, pp. 5466–5470.
- [13] G. H. Golub and C. F. Van Loan, "An Analysis of the Total Least Squares Problem," *SIAM journal on numerical analysis*, vol. 17, no. 6, pp. 883–893, 1980.
- [14] A. Weiss, A. Yeredor, and M. Shtauf, "Iterative Symbol Recovery for Power-Efficient DC-Biased Optical OFDM Systems," *Journal of lightwave technology*, vol. 34, no. 9, pp. 2331–2338, 2016.
- [15] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels," *IEEE Trans. on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.