ASYMPTOTICALLY OPTIMAL RECOVERY OF GAUSSIAN SOURCES FROM NOISY STATIONARY MIXTURES: THE LEAST-NOISY MAXIMALLY-SEPARATING SOLUTION

Amir Weiss and Arie Yeredor

School of Electrical Engineering
Tel-Aviv University
P.O. Box 39040, Tel-Aviv 69978, Israel
{amirwei2@mail, arie@eng}.tau.ac.il

ABSTRACT

We address the problem of source separation from noisy mixtures in a semi-blind scenario, with stationary, temporally-diverse Gaussian sources and known spectra. In such noisy models, a dilemma arises regarding the desired objective. On one hand, a "maximally separating" solution, providing the minimal attainable Interference-to-Source-Ratio (ISR), would often suffer from significant residual noise. On the other hand, optimal Minimum Mean Square Error (MMSE) estimation would yield estimates which are the "least distorted" versions of the true sources, often at the cost of compromised ISR. Based on Maximum Likelihood (ML) estimation of the unknown underlying model parameters, we propose two ML-based estimates of the sources. One asymptotically coincides with the MMSE estimate of the sources, whereas the other asymptotically coincides with the (unbiased) "least-noisy maximally-separating" solution for this model. We prove the asymptotic optimality of the latter and present the corresponding Cramér-Rao lower bound. We discuss the differences in principal properties of the proposed estimates and demonstrate them empirically using simulation results.

Index Terms— Semi-blind source separation, independent component analysis, maximum likelihood, minimum mean square error, least squares.

1. INTRODUCTION

In classical Independent Component Analysis (ICA, [1, 2]) / Blind Source Separation (BSS, [3–5]), the mixtures are assumed to be linear combinations of mutually statistically independent source signals. The term "blind" refers to the fact that no further prior knowledge is available.

However, in some cases, frequently referred to as "semi-blind" [6, 7], some *a-priori* statistical information on the sources is available. A particular case is when the sources' probability distributions are known (possibly up to some unknown parameters), thus enabling the Maximum Likelihood (ML) approach [8–11]. In the noiseless model, the ML Estimate (MLE) of the demixing matrix enjoys the *equivariance* property (e.g., [12, 13]), which essentially means that the resulting Interference-to-Source Ratio (ISR) does not depend on the true value of the mixing matrix but only on the sources' statistics. Unfortunately, this appealing property holds true (for the MLE, as well as for some other separation algorithms) only in the noise-free model, which is possibly the reason why the noisy case has seen less theoretical treatment in the literature than its noiseless counterpart. A few representative examples of separation approaches for noisy models are Joint Approximate Diagonalization of Eigen-matrices (JADE, Cardoso and Souloumiac [14]) for the separation of non-Gaussian sources; an ML approach, based on the expectation-maximization algorithm, for Gaussian mixtures sources (Moulines *et al.* [15]) and the Second-Order Blind Identification (SOBI, Belouchrani *et al.* [16]) algorithm for stationary sources. Among these, only SOBI enables separation of Gaussian sources[1].

In this paper we proceed to explore noisy mixtures of stationary (temporally-diverse) Gaussian sources in a semi-blind scenario, pursuing optimal recovery of the sources in two alternative senses. Based on our recent work [17], we propose two ML-based estimates of the sources: One is asymptotically the Minimum Mean Square Error (MMSE) estimate of the sources, whereas the other is asymptotically the (unbiased) "least-noisy maximally-separating" solution for this model. We compare these estimates and shed some light on the trade-off involved.

1.1. Notations

We use $x, \boldsymbol{x}$ and $\boldsymbol{X}$ for a scalar, column vector and matrix, resp. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{-1}$ denote the transposition and inverse operators, resp. $\mathbb{E}[\cdot]$ denotes expectation. The Kronecker product is denoted by $\otimes$. We also denote by $\boldsymbol{I}_K$ the $K \times K$ identity matrix, and the pinning vector $\boldsymbol{e}_k$ denotes the k-th column of $\boldsymbol{I}_K$. We define $\mathrm{vec}(\cdot)$ as the operator which concatenates the columns of an $M \times N$ matrix into an $MN \times 1$ column vector. The $\mathrm{Diag}(\cdot)$ operator forms an $N \times N$ diagonal matrix from its N-dimensional vector argument. Finally, O is the all zeros-matrix (with proper dimensions).

The first author would like to thank the Yitzhak and Chaya Weinstein Research Institute for Signal Processing for a fellowship.

[1] Under the second-order statistics identifiability condition [16]

2. PROBLEM FORMULATION

Consider the following M sources - L sensors static, instantaneous, linear model

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{V} \in \mathbb{R}^{L\times T}, \tag{1}$$

where $\boldsymbol{S} = [\boldsymbol{s}_1 \ \cdots \ \boldsymbol{s}_M]^{\mathrm{T}} \in \mathbb{R}^{M\times T}$ denotes a matrix of M source signals of length T, $\boldsymbol{A} \in \mathbb{R}^{L\times M}$ is a (deterministic) mixing matrix, $\boldsymbol{V} = [\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_L]^{\mathrm{T}} \in \mathbb{R}^{L\times T}$ denotes a matrix of L additive noise signals (one for each sensor), where we assume $L \geq M$, and the observed mixture signals are given by $\boldsymbol{X} = [\boldsymbol{x}_1 \ \cdots \ \boldsymbol{x}_L]^{\mathrm{T}} \in \mathbb{R}^{L\times T}$. In our semi-blind model, we assume that all the source signals are zero-mean stationary Gaussian processes with known Positive-Definite (PD) Toeplitz temporal covariance matrices $\boldsymbol{C}_s^{(m)} \triangleq \mathbb{E}\left[\boldsymbol{s}_m\boldsymbol{s}_m^{\mathrm{T}}\right]$ (for every $m \in \{1,\dots,M\}$), distinct from one another. As in the standard ICA model, the sources $\boldsymbol{s}_1,\dots,\boldsymbol{s}_M \in \mathbb{R}^{T\times 1}$ (i.e., the rows of $\boldsymbol{S}$) are assumed to be mutually statistically independent and the mixing matrix $\boldsymbol{A}$ is assumed to be unknown. Furthermore, we assume that the noise signals $\boldsymbol{v}_1,\dots,\boldsymbol{v}_L \in \mathbb{R}^{T\times 1}$ from all the sensors (i.e., the rows of $\boldsymbol{V}$) are mutually statistically independent, temporally-white Gaussian noise processes, each with a temporal covariance matrix $\mathbb{E}\left[\boldsymbol{v}_\ell\boldsymbol{v}_\ell^{\mathrm{T}}\right] = \sigma_{v_\ell}^2\boldsymbol{I}_T$ (for every $\ell \in \{1,\dots,L\}$), and are also statistically independent of all the sources. The noises' variances $\sigma_{v_1}^2,\dots,\sigma_{v_L}^2 \in \mathbb{R}^+$ are assumed to be (deterministic) unknown and are denoted collectively in (spatial) matrix form as $\boldsymbol{\Lambda} \triangleq \mathrm{Diag}\left(\left[\sigma_{v_1}^2 \cdots \sigma_{v_L}^2\right]\right) \in \mathbb{R}^{L\times L}$.

Thus, given the measurement matrix $\boldsymbol{X}$ and the sources' covariances $\{\boldsymbol{C}_s^{(m)}\}_{m=1}^M$, our goal is to separate and estimate the unobservable sources $\boldsymbol{s}_1,\dots,\boldsymbol{s}_M$. Note that for this model, in contrary to the classical (fully blind) model, no permutation nor scale ambiguities exist, and the only remaining inevitable ambiguities are sign ambiguities.

As we have recently shown in [17], due to the stationarity and Gaussianity of all the signals in (1), the measurements' distributions, parametrized by the unknown model parameters $\boldsymbol{A},\boldsymbol{\Lambda}$, may be written conveniently by resorting to an equivalent frequency-domain representation of the problem. Consequently, the MLEs of $\boldsymbol{A},\boldsymbol{\Lambda}$, denoted by $\widehat{\boldsymbol{A}}_{\text{ML}},\widehat{\boldsymbol{\Lambda}}_{\text{ML}}$, resp., may be obtained by solving the resulting likelihood equations (eq. (20) in [17]) using, e.g., Fisher's scoring algorithm (see section III-B in [17]). Based on these MLEs, we propose the "least-noisy maximally-separating" solution, presented in the following section.

3. OPTIMAL SOURCES RECOVERY

Assume for the moment that $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ are known, and consider the following equivalent representation of (1)

$$\boldsymbol{x} = (\boldsymbol{A}\otimes\boldsymbol{I}_T)\,\boldsymbol{s} + \boldsymbol{v} \in \mathbb{R}^{LT\times 1}, \tag{2}$$

where $\boldsymbol{x} \triangleq \mathrm{vec}\left(\boldsymbol{X}^{\mathrm{T}}\right) \in \mathbb{R}^{LT\times 1}$, $\boldsymbol{v} \triangleq \mathrm{vec}\left(\boldsymbol{V}^{\mathrm{T}}\right) \in \mathbb{R}^{LT\times 1}$ and $\boldsymbol{s} \triangleq \mathrm{vec}\left(\boldsymbol{S}^{\mathrm{T}}\right) \in \mathbb{R}^{MT\times 1}$.

3.1. Minimum MSE vs. Maximum Separation

In order to consider optimal recovery of the sources $\boldsymbol{s}$ from $\boldsymbol{x}$, unlike in the noiseless case, it is crucial to explicitly define what the desired objective is. One option is to obtain the "closest possible" estimate of the sources, e.g., in the sense of MMSE. In this case, since $\boldsymbol{x}$ and $\boldsymbol{s}$ are jointly Gaussian, the MMSE estimate of $\boldsymbol{s}$ from $\boldsymbol{x}$ (which is also the linear MMSE estimate of $\boldsymbol{s}$ from $\boldsymbol{x}$) is given by

$$\widehat{\boldsymbol{s}}_{\text{MMSE}} \triangleq \mathbb{E}\left[\boldsymbol{s}|\boldsymbol{x}\right] = \boldsymbol{C}_{\boldsymbol{s}\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{x} \in \mathbb{R}^{MT\times 1}, \tag{3}$$

where

$$\boldsymbol{C}_{\boldsymbol{s}\boldsymbol{x}} \triangleq \mathbb{E}\left[\boldsymbol{s}\boldsymbol{x}^{\mathrm{T}}\right] = \boldsymbol{C}_{\boldsymbol{s}}\left(\boldsymbol{A}^{\mathrm{T}}\otimes\boldsymbol{I}_T\right) \in \mathbb{R}^{MT\times LT}, \tag{4}$$

$$\boldsymbol{C}_{\boldsymbol{x}} \triangleq \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right] = (\boldsymbol{A}\otimes\boldsymbol{I}_T)\,\boldsymbol{C}_{\boldsymbol{s}}\left(\boldsymbol{A}^{\mathrm{T}}\otimes\boldsymbol{I}_T\right) + \boldsymbol{\Lambda}\otimes\boldsymbol{I}_T \in \mathbb{R}^{LT\times LT}, \tag{5}$$

$$\boldsymbol{C}_{\boldsymbol{s}} \triangleq \mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\right] = \begin{bmatrix} \boldsymbol{C}_s^{(1)} & \dots & \mathrm{O} \\ \vdots & \ddots & \vdots \\ \mathrm{O} & \dots & \boldsymbol{C}_s^{(M)} \end{bmatrix} \in \mathbb{R}^{MT\times MT}. \tag{6}$$

Accordingly, the m-th source MMSE estimate is given by $(\widehat{\boldsymbol{s}}_m)_{\text{MMSE}} = \widehat{\boldsymbol{S}}_{\text{MMSE}}^{\mathrm{T}}\boldsymbol{e}_m \in \mathbb{R}^{T\times 1}$, where $\widehat{\boldsymbol{s}}_{\text{MMSE}} = \mathrm{vec}\left(\widehat{\boldsymbol{S}}_{\text{MMSE}}^{\mathrm{T}}\right)$. This estimate attains the minimal attainable MSE matrix of *any* estimate, given by

$$\boldsymbol{C}_{\text{MMSE}} \triangleq \mathbb{E}\left[\left(\widehat{\boldsymbol{s}}_{\text{MMSE}} - \boldsymbol{s}\right)\left(\widehat{\boldsymbol{s}}_{\text{MMSE}} - \boldsymbol{s}\right)^{\mathrm{T}}\right] = \boldsymbol{C}_{\boldsymbol{s}} - \boldsymbol{C}_{\boldsymbol{s}\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{s}}. \tag{7}$$

Another possible objective is to obtain "maximal separation" of the sources, even at the cost of a compromised MSE in their estimates. This approach is often termed in the context of communication systems as "zero-forcing" (e.g., [18, 19]) and is known to minimize all (spatial) intersymbol interference. In ICA, a "maximally-separating" solution minimizes the resulting ISR, and in semi-blind scenarios this solution is obtained by applying the (pseudo-) inverse of the MLE of the mixing matrix $\boldsymbol{A}$ to the mixtures' matrix $\boldsymbol{X}$ (as shown in, e.g., [13]). However, if the number of sensors is higher than the number of sources, i.e., $L > M$, then the maximally-separating solution is not unique, and the set of maximally-separating solutions, which would all yield exactly the same ISR, is given by $\widehat{\mathcal{S}}_{\boldsymbol{W}} = \left\{\widehat{\boldsymbol{S}} \in \mathbb{R}^{M\times T} : \widehat{\boldsymbol{S}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\right\}$ and is parametrized by a PD weight matrix $\boldsymbol{W} \in \mathbb{R}^{L\times L}$. Indeed,

$$\forall\widehat{\boldsymbol{S}} \in \widehat{\mathcal{S}}_{\boldsymbol{W}} : \widehat{\boldsymbol{S}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X} = \tag{8}$$

$$\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\left(\boldsymbol{A}\boldsymbol{S} + \boldsymbol{V}\right) = \tag{9}$$

$$\boldsymbol{S} + \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{V} \triangleq \underbrace{\boldsymbol{S}}_{\text{zero ISR}} + \underbrace{\boldsymbol{V}_{\boldsymbol{W}}}_{\text{“}\boldsymbol{W}\text{-colored”}}, \tag{10}$$

regardless of the choice of the weight matrix $\boldsymbol{W}$, where $\boldsymbol{V}_{\boldsymbol{W}}$ is the "$\boldsymbol{W}$-colored" noise. The set $\widehat{\mathcal{S}}_{\boldsymbol{W}}$ is, of course, none other than the Weighted LS (WLS) solutions set, which is obtained by minimizing the WLS criterion at each time-instant when the sources are considered fixed. Nevertheless, one particularly interesting solution from $\widehat{\mathcal{S}}_{\boldsymbol{W}}$ is the Optimally WLS (OWLS) solution, which is obtained by choosing the weight matrix as the inverse of the (spatial) noise covariance matrix, i.e., $\boldsymbol{W} = \boldsymbol{\Lambda}^{-1}$, and is therefore given by

$$\widehat{\boldsymbol{S}}_{\text{OWLS}} \triangleq \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{X} \in \mathbb{R}^{M\times T}. \tag{11}$$

This is in fact the "least-noisy maximally-separating" solution, in the sense that among all maximally-separating solutions $\widehat{\mathcal{S}}_{\boldsymbol{W}}$, (11) attains the minimal noise, such that each column of $\boldsymbol{V}_{\boldsymbol{\Lambda}^{-1}}$ has the minimal attainable covariance matrix by virtue of the Gauss-Markov theorem [20]. Accordingly, (11) is also commonly referred to (in classical Estimation Theory) as the Best Linear Unbiased Estimate (BLUE). Moreover, since in this case the additive noise $\boldsymbol{V}$ is Gaussian, the BLUE is also the MLE of the sources, which is an efficient estimate ([21]) in this problem (even not asymptotically), and therefore is also the Uniformly Minimum-Variance Unbiased Estimate (UMVUE) [22].

3.2. The ML-Based MMSE and OWLS Solutions

Now, recall that $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ are in fact unknown. Therefore, based on similar logic of the ML-based MMSE estimate of the sources (presented in [17]), given by

$$\widehat{\boldsymbol{s}}_{\text{ML-MMSE}} \triangleq \widehat{\boldsymbol{C}}_{\boldsymbol{s}\boldsymbol{x}}\widehat{\boldsymbol{C}}_{\boldsymbol{x}}^{-1}\boldsymbol{x}, \tag{12}$$

where $\widehat{\boldsymbol{C}}_{\boldsymbol{s}\boldsymbol{x}}$ and $\widehat{\boldsymbol{C}}_{\boldsymbol{x}}$ are the MLEs of $\boldsymbol{C}_{\boldsymbol{s}\boldsymbol{x}}$ and $\boldsymbol{C}_{\boldsymbol{x}}$[2], resp., we suggest the ML-based OWLS estimate of the sources

$$\widehat{\boldsymbol{S}}_{\text{ML-OWLS}} \triangleq \left(\widehat{\boldsymbol{A}}_{\text{ML}}^{\mathrm{T}}\widehat{\boldsymbol{\Lambda}}_{\text{ML}}^{-1}\widehat{\boldsymbol{A}}_{\text{ML}}\right)^{-1}\widehat{\boldsymbol{A}}_{\text{ML}}^{\mathrm{T}}\widehat{\boldsymbol{\Lambda}}_{\text{ML}}^{-1}\boldsymbol{X}, \tag{13}$$

based only on the measurements $\boldsymbol{X}$. We stress that for any finite sample size T, $\widehat{\boldsymbol{S}}_{\text{ML-OWLS}} \neq \widehat{\boldsymbol{S}}_{\text{OWLS}}$ almost surely. However, the estimate (13) enjoys an attractive asymptotic optimality property, as we show in the following Theorem.

Theorem 1 (ML-based OWLS convergence to the OWLS) *Let $\widehat{\boldsymbol{A}}_{ML}$, $\widehat{\boldsymbol{\Lambda}}_{ML}$ be the MLEs of $\boldsymbol{A}$, $\boldsymbol{\Lambda}$, resp., and let $\widehat{\boldsymbol{S}}_{ML\text{-}OWLS}$ be the ML-based OWLS estimate of $\boldsymbol{S}$ as defined in* (13). *Then for $T \to \infty$ $\widehat{\boldsymbol{S}}_{ML\text{-}OWLS}$ converges in probability to $\widehat{\boldsymbol{S}}_{OWLS}$.*

Proof 1 *By definition, $\widehat{\boldsymbol{A}}_{ML}$, $\widehat{\boldsymbol{\Lambda}}_{ML}$ are the MLEs of $\boldsymbol{A}$, $\boldsymbol{\Lambda}$, resp. Thus, in particular, they are consistent ([23]). Therefore, from the continuous mapping theorem [24], which states that continuous functions are limit-preserving even if their arguments are sequences of random variables, we have that*

$$\left(\widehat{\boldsymbol{A}}_{ML}, \widehat{\boldsymbol{\Lambda}}_{ML}\right) \xrightarrow[T\to\infty]{p} (\boldsymbol{A}, \boldsymbol{\Lambda}) \Rightarrow \widehat{\boldsymbol{S}}_{ML\text{-}OWLS} \xrightarrow[T\to\infty]{p} \widehat{\boldsymbol{S}}_{OWLS}, \tag{14}$$

since the OWLS is a continuous function of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$. ∎

[2] Obtained by substituting $\widehat{\boldsymbol{A}}_{\text{ML}}, \widehat{\boldsymbol{\Lambda}}_{\text{ML}}$ for $\boldsymbol{A}, \boldsymbol{\Lambda}$, resp., in (4)-(5)

Hence, the estimate (13) asymptotically coincides with the (clairvoyant) least-noisy maximally-separating solution (11), which is the UMVUE, and as T grows, its conditional MSE given the sources decreases and converges (in probability) to

$$\boldsymbol{C}_{\text{CRLB}} \triangleq \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{A}\right)^{-1}, \tag{15}$$

which is the well-known Cramér-Rao Lower Bound (CRLB) on the conditional covariance matrix (given the sources) of any unbiased estimate of the sources, when the nuisance parameters $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ are assumed known. Thus, Theorem 1 implies that (13) is also asymptotically efficient.

We note in passing that one may estimate the bounds $\boldsymbol{C}_{\text{MMSE}}$ and $\boldsymbol{C}_{\text{CRLB}}$ by substituting into (7) and (15), resp., the MLEs $\widehat{\boldsymbol{A}}_{\text{ML}}$ and $\widehat{\boldsymbol{\Lambda}}_{\text{ML}}$. The obtained estimates of the bounds would be their MLEs due to the invariance property of the MLE [25], and therefore would serve as consistent estimates.

3.3. The ML-based MMSE vs. The ML-based OWLS

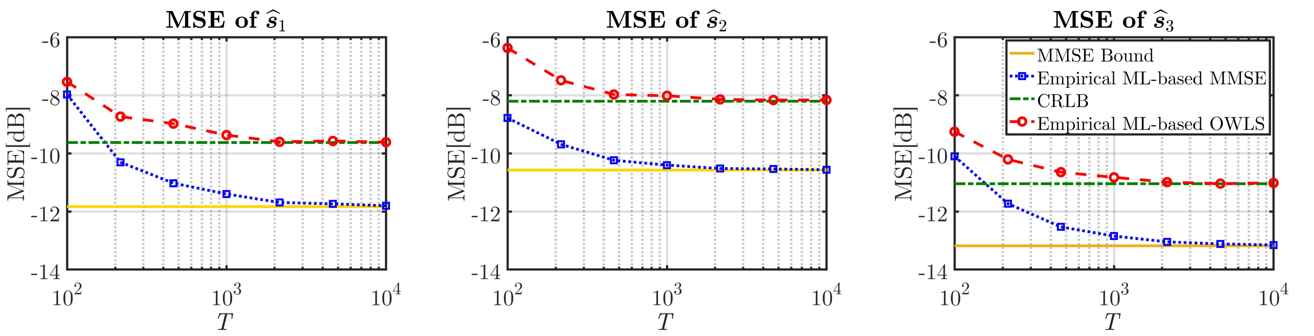
Stemming from two different objectives, although both ML-based estimates (MMSE and OWLS) share asymptotic optimality (w.r.t. different optimality criteria), these estimates differ quite significantly in some important properties:

- **Unbiasedness** - Define a "strongly unbiased" estimate of a source $\boldsymbol{s}_m$ as an estimate whose conditional mean given all the sources is $\boldsymbol{s}_m$ (for $m = 1, \ldots, M$). Then, while the ML-based OWLS is asymptotically strongly unbiased, the ML-based MMSE is strongly biased even when it fully coincides with the (clairvoyant) MMSE estimate (3). Nevertheless, the latter attains the lowest attainable MSE thanks to its bias.

- **Memorylessness** - When the MLEs $\widehat{\boldsymbol{A}}_{\text{ML}}$, $\widehat{\boldsymbol{\Lambda}}_{\text{ML}}$ are considered fixed, the ML-based OWLS is obtained by an instantaneous (memoryless) transformation of the received signals, while the ML-based MMSE applies filtering to the received signals, which is certainly not an instantaneous (memoryless) operation, like the original mixing operation is. Therefore, unlike the ML-based OWLS, the ML-based MMSE estimate distorts the signals with frequency-selective filtering and separation.

- **Consistent Separation** - The ML-based OWLS asymptotically yields zero ISR, meaning perfect separation, and is therefore considered as a consistent separator. Conversely, the ML-based MMSE compromises (consistent) separation for the sake of the minimal total distortion (considering both noise and residual energy from other sources), in the sense of MMSE.

We stress that while (3) and (11), given the nuisance parameters $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$, are both linear estimates of the sources, the ML-based estimates (12) and (13) are certainly not linear estimates, since $\widehat{\boldsymbol{A}}_{\text{ML}}$ and $\widehat{\boldsymbol{\Lambda}}_{\text{ML}}$ are nonlinear functions of the measurements $\boldsymbol{X}$ (as they are solutions of the (nonlinear) likelihood equations, prescribed in [17], for the model (1)).



(a) (b) (c)

Fig. 1: MSE vs. T. Empirical results were obtained by 100 independent trials. (a) MSE of $\widehat{\boldsymbol{s}}_1$ (b) MSE of $\widehat{\boldsymbol{s}}_2$ (c) MSE of $\widehat{\boldsymbol{s}}_3$

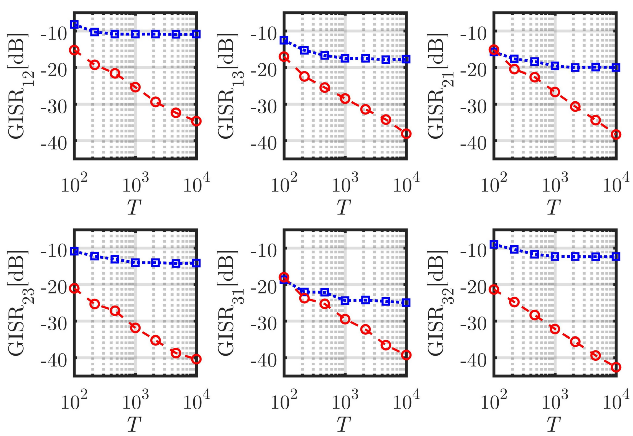


Fig. 2: Average GISR vs. T (using the legend of Fig. 1). Empirical results were obtained by 100 independent trials.

	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{s}_3$
AR(1) Parameter	0.23	0.76	-0.54

Table 1: AR(1) parameters of the sources.

4. SIMULATION RESULTS

We consider a scenario where $\boldsymbol{A} \in \mathbb{R}^{4\times3}$ and the $M = 3$ sources are all Gaussian Auto-Regressive (AR) processes of order 1 (AR(1)), each with unit variance and an AR parameter as presented in Table 1. The elements of the mixing matrix were drawn (once) independently from a standard Normal distribution and the (different) noise variances in each of the sensors were set to $\left[\sigma_{v_1}^2 \ \sigma_{v_2}^2 \ \sigma_{v_3}^2 \ \sigma_{v_4}^2\right] = \left[-10 \ -13\tfrac{1}{3} \ -16\tfrac{2}{3} \ -20\right]$ [dB]. The MLEs of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ were obtained by Fisher's scoring algorithm (as described in detail in [17]). Thereafter, the estimates (12) and (13) were obtained based on these MLEs.

Fig. 1 presents the empirical MSEs of the ML-based MMSE and OWLS estimates, as well as the theoretical lower bounds (the diagonals of $\boldsymbol{C}_{\text{MMSE}}$ and $\boldsymbol{C}_{\text{CRLB}}$), versus the sample size T. As seen in all three figures, the empirical MSEs exhibit a convergence trend towards their corresponding bound, demonstrating the asymptotic optimality of the ML-based estimates. Moreover, the superiority of the ML-based MMSE over the ML-based OWLS, in this sense, is also evident.

Next, we compare the ML-based estimates w.r.t. their separation performance evaluated by the Generalized ISR (GISR), defined as follows. For an estimate of the sources of the form $\widehat{\boldsymbol{s}} = \widetilde{\boldsymbol{B}}\boldsymbol{x} \in \mathbb{R}^{MT\times1}$, where $\widetilde{\boldsymbol{B}} \in \mathbb{R}^{MT\times LT}$, we define the GISR (for all $i \neq j \in \{1,\dots,M\}$) as

$$\text{GISR}_{ij} \triangleq \frac{\text{Tr}\left(\mathbb{E}\left[\widetilde{\boldsymbol{T}}^{(i,j)}\left(\widetilde{\boldsymbol{T}}^{(i,j)}\right)^{\text{T}}\right]\right)}{\text{Tr}\left(\mathbb{E}\left[\widetilde{\boldsymbol{T}}^{(i,i)}\left(\widetilde{\boldsymbol{T}}^{(i,i)}\right)^{\text{T}}\right]\right)} \cdot \frac{\text{Tr}\left(\boldsymbol{C}_s^{(j)}\right)}{\text{Tr}\left(\boldsymbol{C}_s^{(i)}\right)}, \quad (16)$$

where the generalized global demixing-mixing matrix is defined as $\widetilde{\boldsymbol{T}} \triangleq \widetilde{\boldsymbol{B}}\left(\boldsymbol{A}\otimes\boldsymbol{I}_T\right)$, denoting its block-partition as

$$\widetilde{\boldsymbol{T}} \triangleq \begin{bmatrix} \widetilde{\boldsymbol{T}}^{(1,1)} & \dots & \widetilde{\boldsymbol{T}}^{(1,M)} \\ \vdots & \ddots & \vdots \\ \widetilde{\boldsymbol{T}}^{(M,1)} & \dots & \widetilde{\boldsymbol{T}}^{(M,M)} \end{bmatrix} \in \mathbb{R}^{MT\times MT}. \quad (17)$$

Note that while the classical definition of ISR (e.g., [13]) is restricted to instantaneous (memoryless) separation, the GISR extends the same notion to general (not necessarily instantaneous) linear separators, and coincides with ISR for instantaneous separation.

Fig. 2 presents the resulting GISR elements of both ML-based estimates (12) and (13). Here, the superiority of the ML-based OWLS over the ML-based MMSE (in this sense) is demonstrated, in addition to its separation consistency property, which is reflected by the (asymptotically) constant rate of decay of the GISR, as expected.

5. CONCLUSION

In the context of semi-blind separation of temporally-diverse Gaussian sources, we contrasted the ML-based MMSE and OWLS estimates of sources. While compromising the overall MSE, the ML-based OWLS estimate was shown (both analytically and empirically) to be asymptotically efficient, attaining the CRLB on the MSE of any unbiased estimate of the sources, while enjoying separation consistency.

We note that the proposed estimates may also be used for non-Gaussian sources, based on Gaussian Quasi ML (QML) estimation, which was shown in [17] to be consistent. In such cases the QML-based OWLS estimate would retain its asymptotic efficiency.

6. REFERENCES

[1] Pierre Comon and Christian Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.

[2] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent component analysis*, vol. 46, John Wiley & Sons, 2004.

[3] Jean-François Cardoso, "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem," in *Proc. of ICASSP*, 1990, pp. 2655–2658.

[4] Christian Jutten and Jeanny Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal processing*, vol. 24, no. 1, pp. 1–10, 1991.

[5] Pierre Comon, Christian Jutten, and Jeanny Herault, "Blind separation of sources, Part II: Problems statement," *Signal processing*, vol. 24, no. 1, pp. 11–20, 1991.

[6] Jake Gunther, "Learning echo paths during continuous double-talk using semi-blind source separation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 2, pp. 646–660, 2012.

[7] Francesco Nesta, Ted S. Wada, and Biing-Hwang Juang, "Batch-online semi-blind source separation applied to multi-channel acoustic echo cancellation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 3, pp. 583–599, 2011.

[8] Dinh Tuan Pham and Philippe Garat, "Blind Separation of Mixture of Independent Sources Through a Maximum Likelihood Approach," in *Proc. of EUSIPCO*, 1997, pp. 771–774.

[9] Jean-François Cardoso, "Infomax and maximum likelihood for blind source separation," *IEEE Signal Processing Letters*, vol. 4, no. 4, pp. 112–114, 1997.

[10] Serge Dégerine and Abdelhamid Zaïdi, "Separation of an instantaneous mixture of Gaussian autoregressive sources by the exact maximum likelihood approach," *IEEE Trans. on Signal Processing*, vol. 52, no. 6, pp. 1499–1512, 2004.

[11] Arie Yeredor, "Empirical weighting for blind source separation in a multiple-snapshots scenario," in *Proc. of ICASSP*, 2011, pp. 3704–3707.

[12] Jean-François Cardoso and Beate Hvam Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, 1996.

[13] Arie Yeredor, "Blind separation of Gaussian sources with general covariance structures: Bounds and optimal estimation," *IEEE Trans. on Signal Processing*, vol. 58, no. 10, pp. 5057–5068, 2010.

[14] Jean-François Cardoso and Antoine Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE proceedings F (radar and signal processing)*. IET, 1993, vol. 140, pp. 362–370.

[15] Eric Moulines, Jean-François Cardoso, and Elisabeth Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *Proc. of ICASSP*, 1997, vol. 5, pp. 3617–3620.

[16] Adel Belouchrani, Karim Abed-Meraim, Jean-François Cardoso, and Eric Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[17] Amir Weiss and Arie Yeredor, "A Maximum Likelihood-Based Minimum Mean Square Error Separation and Estimation of Stationary Gaussian Sources from Noisy Mixtures," submitted to *IEEE Trans. on Signal Processing*; also: *ArXiv e-prints*, Oct. 2018, arXiv:1810.09165 [stat.AP], https://arxiv.org/abs/1810.09165.

[18] Anja Klein, Ghassan Kawas Kaleh, and Paul Walter Baier, "Zero forcing and minimum mean-square-error equalization for multiuser detection in code-division multiple-access channels," *IEEE Trans. on Vehicular Technology*, vol. 45, no. 2, pp. 276–287, 1996.

[19] Ami Wiesel, Yonina C. Eldar, and Shlomo Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. on Signal Processing*, vol. 56, no. 9, pp. 4409–4418, 2008.

[20] John Johnston and John DiNardo, *Econometric methods*, vol. 2, New York, 1972.

[21] Ronald A. Fisher, "On the mathematical foundations of theoretical statistics," *Phil. Trans. R. Soc. Lond. A*, vol. 222, no. 594-604, pp. 309–368, 1922.

[22] Erich L. Lehmann and George Casella, *Theory of point estimation*, Springer Science & Business Media, 2006.

[23] Harald Cramér, *Mathematical Methods Of Statistics*, Princeton University Press, 1946.

[24] Henry B. Mann and Abraham Wald, "On stochastic limit and order relationships," *The Annals of Mathematical Statistics*, vol. 14, no. 3, pp. 217–226, 1943.

[25] Nitis Mukhopadhyay, *Probability and statistical inference*, CRC Press, 2000.