Blind Separation of Gaussian Sources With General Covariance Structures: Bounds and Optimal Estimation

Arie Yeredor, *Senior Member, IEEE*

***Abstract*—We consider the separation of Gaussian sources exhibiting general, arbitrary (not necessarily stationary) covariance structures. First, assuming a semi-blind scenario, in which the sources' covariance structures are known, we derive the maximum likelihood estimate of the separation matrix, as well as the induced Cramér–Rao lower bound (iCRLB) on the attainable Interference to Source Ratio (ISR). We then extend our results to the fully blind scenario, in which the covariance structures are unknown. We show that (under a scaling convention) the Fisher information matrix in this case is block-diagonal, implying that the same iCRLB (as in the semi-blind scenario) applies in this case as well. Subsequently, we demonstrate that the same "semi-blind" optimal performance can be approached asymptotically in the "fully blind" scenario if the sources are sufficiently ergodic, or if multiple snapshots are available.**

***Index Terms*—Blind source separation, independent component analysis, nonstationarity, second-order statistics, time-varying AR processes.**

I. Introduction

BLIND source separation (BSS) consists of recovering unobserved source signals from their observed mixtures. In the classical BSS setup the mixture is assumed to be real-valued, linear, static, square-invertible and noiseless, and can be expressed using the following matrix notation:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} \tag{1}$$

where $\boldsymbol{S} \triangleq [\boldsymbol{s}_1\, \boldsymbol{s}_2 \cdots \boldsymbol{s}_K]^T$ is a $K \times N$ matrix containing the K unobserved source signals (each of length N) as its rows; $\boldsymbol{A}$ is the unknown $K \times K$ mixing matrix (assumed to be nonsingular); and $\boldsymbol{X} \triangleq [\boldsymbol{x}_1\, \boldsymbol{x}_2 \cdots \boldsymbol{x}_K]^T$ is the $K \times N$ matrix of K observed mixture signals. The sources (and therefore the observations) are all assumed to have zero mean. We shall denote the $K \times K$ demixing matrix as $\boldsymbol{B} \triangleq \boldsymbol{A}^{-1}$.

The term "blind" usually implies that the mixing matrix $\boldsymbol{A}$ is completely unknown and that the only available information regarding the sources is their mutual statistical independence (giving rise to the term independent component analysis (ICA) in this context). In a "semi-blind" scenario, more *a priori* structural or statistical information about the mixing matrix and/or about the sources might be available. In the context of this work we shall use the term "semi-blind" to refer to cases in which the additional information pertains only to statistical properties of the individual sources.

Generally, second-order statistics (SOS) are insufficient for solving the BSS problem (i.e., for obtaining consistent estimates of the sources), neither in a blind nor in our semi-blind scenario. For example, when each of the sources has an independent, identically distributed (i.i.d.) time-structure, even perfect knowledge of the observations' SOS can only be used for spatial whitening, leaving an unknown residual (orthogonal) mixing, which can only be resolved using some additional statistics, such as higher-order statistics (HOS). Consequently, it is well-known that Gaussian sources with i.i.d. time-structures (or with otherwise similar temporal covariance structures) cannot be separated, since the mixtures' HOS are invariant with respect to (w.r.t.) the mixing (as they all vanish).

Nevertheless, when the sources have different temporal covariance structures (e.g., when they are stationary with distinct spectra), consistent separation can rely exclusively on SOS. Consequently, Gaussian sources can be separated in such cases. Moreover, the ability to work within the confines of a Gaussian model (clearly a widely adopted framework in statistical signal processing, which is nonetheless despicable in classical ICA with i.i.d. sources) has several advantages in this context: It enables a tractable derivation of maximum likelihood (ML) separation, as well as of the induced Cramér–Rao lower bound (iCRLB) on the attainable Interference to source ratio (ISR).[1] The ML separation in this context is asymptotically efficient, attaining the iCRLB.

Prior work on ML separation and on bounds for Gaussian sources has been limited to a few, very important yet merely particular cases in terms of the sources' temporal covariance structure. The most prominent is the case of stationary sources: The "quasi-ML" (QML) approach [18], proposed by Pham and Garat in 1997, is based on expressing the stationary observations' joint probability density function (pdf) in the frequency-domain, assuming arbitrary sources' spectra. When the presumed spectra are the true spectra of the sources (e.g., in a "semi-blind" scenario), QML becomes ML (for Gaussian sources). The "exact ML" (EML) approach [8], proposed by Dégerine and Zaïdi in 2004, further assumes that the sources

Manuscript received October 20, 2009; accepted May 31, 2010. Date of publication June 21, 2010; date of current version September 15, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ljubisa Stankovic.

The author is with the School of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel (e-mail: arie@eng.tau.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2053362

[1]We use the term "induced", since the CRLB does not address the ISR directly, but rather bounds the variance in unbiased estimation of the mixing (or demixing) matrix' elements, which in turn implies an "induced" bound on the ISR—see further discussion in Section II

can be modeled as Gaussian auto-regressive (AR) processes, and is based on expressing the observations' pdf in time-domain, maximizing w.r.t. the unknown mixing matrix and AR parameters simultaneously. Other approaches, which, although not based directly on ML, can be shown to coincide (asymptotically, in the respective Gaussian models) with QML and EML, are Pham's Gaussian mutual information (GMI) approach [14], and the weights-adjusted second-order blind identification (WASOBI) approach for AR sources, proposed by Yeredor [26] and by Tichavský *et al.* [21], [23]. With their EML derivation in [8], Dégerine and Zaïdi further provide an expression for the CRLB on estimation of the demixing matrix' elements. In [9] this result is translated into the induced bound (iCRLB) on the attainable ISR (for Gaussian AR, moving average (MA) and ARMA sources). The derivation of the QML approach in [18] also includes asymptotic performance analysis, which essentially yields an expression for the iCRLB for the case of Gaussian stationary processes with general spectra.

Another particular case (in terms of the sources' temporal covariance structure), involving non-stationary structures, was considered by Pham and Cardoso in [17], where ML separation was applied to Gaussian block-i.i.d. sources. The block-i.i.d. model assumes that the observation interval can be divided into several blocks, such that each source has an i.i.d. time-structure within each block (with different, unknown variances). A slightly more general block-AR model (in which the sources have different, unknown AR time-structures within blocks) was considered by Tichavský *et al.* in [24], where the iCRLB for this model was provided together with a separation scheme which is asymptotically equivalent to ML.

Wide and general as the stationary model may be, in terms of the sources' temporal covariance matrices it is still merely a particular case, in which these matrices have a special (Toeplitz) structure. Naturally, so are the block-i.i.d. (block-diagonal with constant-diagonal blocks) and block-AR (block-diagonal with Toeplitz blocks) models. Evidently, in many cases of interest some or all of the sources may exhibit far more rich and diverse nonstationary structures, e.g., time-varying AR (TVAR) patterns, cyclo-stationary patterns, transient patterns, or any other general (arbitrary) covariance structure. However, to the best of our knowledge, no general framework for ML separation (or for the associated bounds) of Gaussian sources with arbitrary covariance structures has been presented so far. It is our purpose in this paper to fill this gap.

Rather than consider additional particular cases, we shall assume that each source $\boldsymbol{s}_k$ has a *general* $(N \times N)$ covariance matrix $\boldsymbol{C}_k \triangleq E[\boldsymbol{s}_k\boldsymbol{s}_k^T]$, $k = 1,\ldots,K$, not necessarily structured or constrained in any way. In the first part of the paper (in Section III) we shall assume a "semi-blind" scenario, in which all $\boldsymbol{C}_k$ $(k = 1,\ldots K)$ are known, deriving the ML estimate of the demixing (or mixing) matrix, as well as the iCRLB for this case. Then, in the second part (in Section IV), we shall relax this (often unrealistic) assumption, and consider the "fully blind" scenario, in which these matrices are unknown. We shall show that under some scaling constraints, the Fisher information matrix (FIM) in this scenario is block-diagonal with separate blocks corresponding to the mixing parameters and to the unknown covariance parameters, implying that the same iCRLB from the "semi-blind" scenario is also applicable in the blind scenario. Accordingly, we shall also demonstrate (by simulation, in Section V) that with some models, the unknown covariance matrices can be consistently estimated from the observed mixtures (e.g., when the sources are sufficiently ergodic, or when several snapshots are available), leading to asymptotically efficient ML separation (attaining the iCRLB) in the fully blind scenario as well. We begin (in the next section) by outlining some general properties of the iCRLB, which will be useful in the subsequent derivations.

II. The Induced Cramér–Rao Bound on the ISR

The CRLB is a well-known lower-bound on the mean-square estimation error (MSE) of any unbiased estimate of a (deterministic) parameters vector. In the context of BSS, the unknown parameters vector consists of elements of the mixing (or demixing) matrix, and, in a fully blind scenario, also of parameters related to the unknown sources' distributions.[2] The estimated demixing matrix (or the inverse of the estimated mixing matrix) is applied to the observed data in order to recover the sources. Any errors in the estimation of the mixing or demixing matrix would be reflected in some residual mixing in the recovered sources.

Denoting the estimated demixing matrix (based on the observed signal-matrix $\boldsymbol{X}$) as $\hat{\boldsymbol{B}}(\boldsymbol{X})$, we define $\boldsymbol{T}(\boldsymbol{X}) \triangleq \hat{\boldsymbol{B}}(\boldsymbol{X})\boldsymbol{A}$ as the overall mixing-unmixing matrix. The resulting ISR can then be described by a $K \times K$ matrix whose (k,ℓ)th element is given by

$$\mathrm{ISR}_{k,\ell} \triangleq E[T_{k,\ell}^2(\boldsymbol{X})] \cdot \frac{E[\boldsymbol{s}_\ell^T\boldsymbol{s}_\ell]}{E[\boldsymbol{s}_k^T\boldsymbol{s}_k]} \quad 1 \le k \ne \ell \le K \quad (2)$$

(where $T_{k,\ell}(\boldsymbol{X})$ denotes the (k,ℓ)th element of $\boldsymbol{T}(\boldsymbol{X})$), such that this value represents the relative residual energy of the ℓth source in the reconstruction of the kth source.

Noting that $\boldsymbol{T}(\boldsymbol{X})$ is a linear function of $\hat{\boldsymbol{B}}(\boldsymbol{X})$, it is evident that any lower-bound on the MSE in the estimation of $\boldsymbol{B}$ induces an easily-obtained element-wise lower-bound on the attainable ISR matrix. Thus, the CRLB on the MSE in unbiased estimation of $\boldsymbol{B}$ induces a bound on the ISR, which we term the "induced CRLB" (iCRLB). The iCRLB is a linear transformation of the CRLB on $\boldsymbol{B}$. A basic property of the iCRLB, which is instrumental in facilitating our subsequent derivations, is its *equivarance*.

The property of equivariance of ICA algorithms has been long observed and advocated by Cardoso and others (e.g., [5]–[7], [13]): Evidently, any estimator $\hat{\boldsymbol{B}}(\boldsymbol{X})$ of $\boldsymbol{B}$ satisfying $\hat{\boldsymbol{B}}(\boldsymbol{Q}\boldsymbol{X}) = \hat{\boldsymbol{B}}(\boldsymbol{X}) \cdot \boldsymbol{Q}^{-1}$ for all nonsingular $\boldsymbol{Q}$ (and all $\boldsymbol{X}$) would result in

$$\boldsymbol{T}(\boldsymbol{X}) = \hat{\boldsymbol{B}}(\boldsymbol{X})\boldsymbol{A} = \hat{\boldsymbol{B}}(\boldsymbol{A}\boldsymbol{S})\boldsymbol{A} = \hat{\boldsymbol{B}}(\boldsymbol{S})\boldsymbol{A}^{-1}\boldsymbol{A} = \hat{\boldsymbol{B}}(\boldsymbol{S}) \quad (3)$$

which is independent of (or "equivariant in") the true value of the mixing-matrix $\boldsymbol{A}$, and depends only on the sources' realization $\boldsymbol{S}$. Hence, for such estimators, which we term "ISR-equivariant estimators," the resulting ISR would depend only on the statistics of the sources matrix $\boldsymbol{S}$, but not on $\boldsymbol{A}$. The equivariance property is shared by estimators obtained by many (but certainly not by all) classical ICA algorithms.

It is therefore intuitively plausible (yet not obvious in general) that the iCRLB should exhibit a similar invariance property. In the Appendix, we prove the invariance of the iCRLB

[2]Possibly involving continuous distributions in a "semi-parametric" framework—see, e.g., [3].

(for both the semi-blind and fully blind scenarios) by exploiting the asymptotic efficiency of the ML estimate, combined with its own equivariance.

This property is not only conceptually, but also practically significant: It allows to obtain the iCRLB by calculating the CRLB for a specific, conveniently chosen value of the mixing-matrix—knowing that the same ISR bound would be obtained with any (nonsingular) mixing matrix. It is important to realize here, that only the iCRLB—but not the CRLB itself—is invariant in $\boldsymbol{A}$. Therefore, the ability to use the calculation of the CRLB for just one particular choice of $\boldsymbol{A}$ in order to obtain the general iCRLB (for all $\boldsymbol{A}$) is of considerable advantage.

We now turn to the explicit derivation of the iCRLB and ML estimate, concentrating on the semi-blind case first.

III. The Semi-Blind Case

Let us use a somewhat uncommon formulation to rewrite the mixing model (1) as

$$\boldsymbol{x} = (\boldsymbol{A} \otimes \boldsymbol{I}_N)\boldsymbol{s} \tag{4}$$

where $\boldsymbol{s}$ (respectively, $\boldsymbol{x}$) denotes a $KN \times 1$ vector, formed by the concatenation of all source (respectively, observation) signals,

$$\begin{aligned}\boldsymbol{s} &= [\boldsymbol{s}_1^T \cdots \boldsymbol{s}_K^T]^T = \mathrm{vec}(\boldsymbol{S}^T)\\ \boldsymbol{x} &= [\boldsymbol{x}_1^T \cdots \boldsymbol{x}_K^T]^T = \mathrm{vec}(\boldsymbol{X}^T)\end{aligned} \tag{5}$$

$\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix, and $\otimes$ denotes Kronecker's product

$$\boldsymbol{P} \otimes \boldsymbol{Q} = \begin{bmatrix} P_{1,1}\boldsymbol{Q} & P_{1,2}\boldsymbol{Q} & \cdots \\ P_{2,1}\boldsymbol{Q} & P_{2,2}\boldsymbol{Q} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \tag{6}$$

with $P_{k,\ell}$ denoting the (k,ℓ)th element of $\boldsymbol{P}$.

According to the model assumptions, $\boldsymbol{s}$, and therefore also $\boldsymbol{x}$, are zero-mean random vectors. The $KN \times KN$ covariance matrix $\boldsymbol{C_s}$ of $\boldsymbol{s}$ is a block-diagonal matrix composed of $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_K$ as its blocks

$$\boldsymbol{C_s} = E[\boldsymbol{s}\boldsymbol{s}^T] = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{C}_K \end{bmatrix}. \tag{7}$$

The $KN \times KN$ covariance matrix $\boldsymbol{C_x}$ of $\boldsymbol{x}$ is therefore given by

$$\begin{aligned}\boldsymbol{C_x} &= (\boldsymbol{A} \otimes \boldsymbol{I}_N)\boldsymbol{C_s}(\boldsymbol{A} \otimes \boldsymbol{I}_N)^T \\ &= (\boldsymbol{B}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{C_s}(\boldsymbol{B}^{-T} \otimes \boldsymbol{I}_N)\end{aligned} \tag{8}$$

(where $\boldsymbol{B}^{-T}$ is shorthand for $(\boldsymbol{B}^T)^{-1}$), and its inverse is given by

$$\boldsymbol{C_x}^{-1} = (\boldsymbol{B}^T \otimes \boldsymbol{I}_N)\boldsymbol{C_s}^{-1}(\boldsymbol{B} \otimes \boldsymbol{I}_N). \tag{9}$$

Assuming, in addition, that all sources are Gaussian and mutually statistically independent, it follows that the observations vector $\boldsymbol{x}$ is also a (zero-mean) Gaussian vector, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C_x})$.

We shall now derive the CRLB on unbiased estimation of $\boldsymbol{B}$, assuming, in accordance with the semi-blind scenario, that all $\boldsymbol{C}_k$ $(k = 1, \ldots, K)$ are known. This bound will in turn be used for obtaining the iCRLB for this problem. In addition, we shall derive the ML estimate of $\boldsymbol{B}$.

A. The CRLB and iCRLB

Since all the distribution-related parameters (the $\boldsymbol{C}_k$ matrices) are known, the vector of unknown parameters consists only of elements of the unknown mixing matrix, $\boldsymbol{\xi} = \mathrm{vec}(\boldsymbol{B})$. Due to the zero-mean Gaussian distribution of $\boldsymbol{x}$, the respective elements of the $K^2 \times K^2$ FIM for the (k,ℓ)th and (p,q)th elements of the mixing matrix are well-known to be given by (see, e.g., [11])

$$\begin{aligned}J_{B_{k,\ell},B_{p,q}}(\boldsymbol{A}) &= \frac{1}{2}\mathrm{Tr}\left\{\frac{\partial \boldsymbol{C_x}}{\partial B_{k,\ell}} \cdot \boldsymbol{C_x}^{-1} \cdot \frac{\partial \boldsymbol{C_x}}{\partial B_{p,q}} \cdot \boldsymbol{C_x}^{-1}\right\} \\ &= \frac{1}{2}\mathrm{Tr}\left\{\frac{\partial \boldsymbol{C_x}^{-1}}{\partial B_{k,\ell}} \cdot \boldsymbol{C_x} \cdot \frac{\partial \boldsymbol{C_x}^{-1}}{\partial B_{p,q}} \cdot \boldsymbol{C_x}\right\}\end{aligned} \tag{10}$$

where the second expression follows by substituting the identity

$$\frac{\partial \boldsymbol{P}(\theta)}{\partial \theta} = -\boldsymbol{P}(\theta) \cdot \frac{\partial \boldsymbol{P}^{-1}(\theta)}{\partial \theta} \cdot \boldsymbol{P}(\theta) \tag{11}$$

(which holds for any invertible matrix $\boldsymbol{P}(\theta)$ which depends on a parameter θ), and by using $\mathrm{Tr}\{\boldsymbol{PQ}\} = \mathrm{Tr}\{\boldsymbol{QP}\}$. Let $\boldsymbol{e}_k$ denote the kth column of the $K \times K$ identity matrix, and let $\boldsymbol{E}_{k,\ell} \triangleq \boldsymbol{e}_k\boldsymbol{e}_\ell^T$ denote the $K \times K$ matrix which is all-zeros except for a 1 at the (k,ℓ)th element. Evidently $\partial \boldsymbol{B}/\partial B_{k,\ell} = \boldsymbol{E}_{k,\ell}$. We therefore have from (9), using the relation $(\boldsymbol{C} \otimes \boldsymbol{D})(\boldsymbol{P} \otimes \boldsymbol{Q}) = (\boldsymbol{CP}) \otimes (\boldsymbol{DQ})$ and exploiting the block-diagonality of $\boldsymbol{C_s}^{-1}$,

$$\begin{aligned}\frac{\partial \boldsymbol{C_x}^{-1}}{\partial B_{k,\ell}} &= (\boldsymbol{E}_{\ell,k} \otimes \boldsymbol{I}_N)\boldsymbol{C_s}^{-1}(\boldsymbol{B} \otimes \boldsymbol{I}_N)^T \\ &\quad + (\boldsymbol{B}^T \otimes \boldsymbol{I}_N)\boldsymbol{C_s}^{-1}(\boldsymbol{E}_{k,\ell} \otimes \boldsymbol{I}_N) \\ &= (\boldsymbol{E}_{\ell,k} \otimes \boldsymbol{C}_k^{-1})(\boldsymbol{B} \otimes \boldsymbol{I}_N)^T \\ &\quad + (\boldsymbol{B}^T \otimes \boldsymbol{I}_N)(\boldsymbol{E}_{k,\ell} \otimes \boldsymbol{C}_k^{-1}) \\ &= (\boldsymbol{E}_{\ell,k}\boldsymbol{B}) \otimes \boldsymbol{C}_k^{-1} + (\boldsymbol{B}^T\boldsymbol{E}_{k,\ell}) \otimes \boldsymbol{C}_k^{-1} \\ &= [\boldsymbol{E}_{\ell,k}\boldsymbol{B} + \boldsymbol{B}^T\boldsymbol{E}_{k,\ell}] \otimes \boldsymbol{C}_k^{-1}.\end{aligned} \tag{12}$$

Multiplying on the right with $\boldsymbol{C_x}$, we get

$$\begin{aligned}\frac{\partial \boldsymbol{C_x}^{-1}}{\partial B_{k,\ell}} \cdot \boldsymbol{C_x} &= \left([\boldsymbol{E}_{\ell,k}\boldsymbol{B} + \boldsymbol{B}^T\boldsymbol{E}_{k,\ell}] \otimes \boldsymbol{C}_k^{-1}\right) \\ &\quad \times \left(\boldsymbol{B}^{-1} \otimes \boldsymbol{I}_N\right)\boldsymbol{C}_s\left(\boldsymbol{B}^{-T} \otimes \boldsymbol{I}_N\right) \\ &= \left([\boldsymbol{E}_{\ell,k} + \boldsymbol{B}^T\boldsymbol{E}_{k,\ell}\boldsymbol{B}^{-1}] \otimes \boldsymbol{C}_k^{-1}\right) \\ &\quad \times \boldsymbol{C}_s\left(\boldsymbol{B}^{-T} \otimes \boldsymbol{I}_N\right).\end{aligned} \tag{13}$$

Recall now, that due to the invariance of the iCRLB w.r.t. $\boldsymbol{A}$, we may obtain the iCRLB from the derivation of the CRLB with *any* $\boldsymbol{A}$ (or $\boldsymbol{B}$). For convenience, we now choose to proceed with $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{I}$ (the $K \times K$ identity matrix). Substituting into (13), we get

$$\begin{aligned}\left.\frac{\partial \boldsymbol{C_x}^{-1}}{\partial B_{k,\ell}} \cdot \boldsymbol{C_x}\right|_{\boldsymbol{B}=\boldsymbol{I}} &= \left([\boldsymbol{E}_{\ell,k} + \boldsymbol{E}_{k,\ell}] \otimes \boldsymbol{C}_k^{-1}\right)\boldsymbol{C}_s \\ &= \left[\boldsymbol{E}_{\ell,k} \otimes \boldsymbol{I}_N + \boldsymbol{E}_{k,\ell} \otimes (\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell)\right]\end{aligned} \tag{14}$$

where we have used the relation $[\boldsymbol{E}_{m,n} \otimes \boldsymbol{C}_k^{-1}]\boldsymbol{C_S} = \boldsymbol{E}_{m,n} \otimes (\boldsymbol{C}_k^{-1}\boldsymbol{C}_n)$. Substituting into the FIM (for $B_{k,\ell}$ and for $B_{p,q}$), we get

$$\begin{aligned} & J_{B_{k,\ell},B_{p,q}}(\boldsymbol{I}) \\ & \quad = \frac{1}{2}\mathrm{Tr}\left\{\left[\boldsymbol{E}_{\ell,k} \otimes \boldsymbol{I}_N + \boldsymbol{E}_{k,\ell} \otimes (\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell)\right]\right. \\ & \qquad\qquad \left.\left[\boldsymbol{E}_{q,p} \otimes \boldsymbol{I}_N + \boldsymbol{E}_{p,q} \otimes (\boldsymbol{C}_p^{-1}\boldsymbol{C}_q)\right]\right\} \\ & = \frac{1}{2}\mathrm{Tr}\left\{(\boldsymbol{E}_{\ell,k}\boldsymbol{E}_{q,p}) \otimes \boldsymbol{I}_N + (\boldsymbol{E}_{\ell,k}\boldsymbol{E}_{p,q})\right. \\ & \qquad \otimes(\boldsymbol{C}_p^{-1}\boldsymbol{C}_q) + (\boldsymbol{E}_{k,\ell}\boldsymbol{E}_{q,p}) \otimes (\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell) \\ & \qquad \left. +(\boldsymbol{E}_{k,\ell}\boldsymbol{E}_{p,q}) \otimes (\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell\boldsymbol{C}_p^{-1}\boldsymbol{C}_q)\right\}. \end{aligned} \tag{15}$$

Substituting the general relation

$$\begin{aligned} \mathrm{Tr}\{(\boldsymbol{E}_{i,j}\boldsymbol{E}_{m,n}) \otimes \boldsymbol{P}\} &= \mathrm{Tr}\{(\boldsymbol{e}_i\boldsymbol{e}_j^T\boldsymbol{e}_m\boldsymbol{e}_n^T) \otimes \boldsymbol{P}\} \\ &= \delta_{jm}\mathrm{Tr}\{\boldsymbol{E}_{i,n} \otimes \boldsymbol{P}\} \\ &= \delta_{jm}\delta_{in}\mathrm{Tr}\{\boldsymbol{P}\} \end{aligned} \tag{16}$$

(where δ_{jm} denotes Kronecker's delta, which equals 1 if $j = m$ and 0 if $j \neq m$), we get

$$\begin{aligned} & J_{B_{k,\ell},B_{p,q}}(\boldsymbol{I}) \\ & = \frac{1}{2}\left[\delta_{kq}\delta_{\ell p}\left(N + \mathrm{Tr}\left\{\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell\boldsymbol{C}_p^{-1}\boldsymbol{C}_q\right\}\right)\right. \\ & \qquad \left. +\delta_{kp}\delta_{\ell q}\left(\mathrm{Tr}\left\{\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell\right\} + \mathrm{Tr}\left\{\boldsymbol{C}_p^{-1}\boldsymbol{C}_q\right\}\right)\right] \\ & = \frac{1}{2}\left[\delta_{kq}\delta_{\ell p}(N+N)+\delta_{kp}\delta_{\ell q}\left(\mathrm{Tr}\{\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell\}+\mathrm{Tr}\{\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell\}\right)\right] \\ & = \delta_{kq}\delta_{\ell p} \cdot N + \delta_{kp}\delta_{\ell q}\mathrm{Tr}\left\{\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell\right\}. \end{aligned} \tag{17}$$

Put differently, this expression reads

$$J_{B_{k,\ell},B_{p,q}}(\boldsymbol{I}) = \begin{cases} \mathrm{Tr}\left\{\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell\right\} & (k,\ell) = (p,q),\ k \neq \ell \\ N & (k,\ell) = (q,p),\ k \neq \ell \\ 2N & k = \ell = p = q \\ 0 & \text{elsewhere} \end{cases} \tag{18}$$

which means that the $K^2 \times K^2$ FIM $\boldsymbol{J}(\boldsymbol{I})$ is essentially block-diagonal, with $K(K-1)/2$ blocks (each 2×2) accounting for elements-couples of the form $(k,\ell),(k,\ell)$ and $(k,\ell),(\ell,k)$ for $k \neq \ell$; and additional K diagonal elements accounting for all (k,k) terms. Consequently, the corresponding CRLB matrix is a similarly-structured, essentially block-diagonal matrix, with the respective 2×2 inverses substituting the 2×2 blocks, and $1/2N$ substituting the diagonal elements. Thus, defining

$$\phi_{k,\ell} \triangleq \frac{1}{N}\mathrm{Tr}\left\{\boldsymbol{C}_k\boldsymbol{C}_\ell^{-1}\right\} = \frac{1}{N}\mathrm{Tr}\left\{\boldsymbol{C}_\ell^{-1}\boldsymbol{C}_k\right\} \tag{19}$$

we may express the CRLB on unbiased estimation of elements of $\boldsymbol{B}$ when $\boldsymbol{B} = \boldsymbol{I}$ as follows:

$$\begin{aligned} \mathrm{cov}\left(\begin{bmatrix}\hat{B}_{k,\ell} \\ \hat{B}_{\ell,k}\end{bmatrix}\right) &\geq \frac{1}{N}\begin{bmatrix}\phi_{\ell,k} & 1 \\ 1 & \phi_{k,\ell}\end{bmatrix}^{-1}, \\ \mathrm{var}\left(\hat{B}_{kk}\right) &\geq \frac{1}{2N} \end{aligned} \tag{20}$$

for $0 \leq k \neq \ell \leq K$ (with differently-indexed cross-terms being zero). In particular, this means that the estimation variances of all $B_{k,\ell}$ (with $k \neq \ell$) are bounded by

$$\mathrm{var}\left(\hat{B}_{k,\ell}\right) \geq \frac{1}{N} \cdot \frac{\phi_{k,\ell}}{\phi_{k,\ell}\phi_{\ell,k} - 1}. \tag{21}$$

In order to obtain the iCRLB, we may now use the relation

$$\boldsymbol{T}(\boldsymbol{X}) = \hat{\boldsymbol{B}}(\boldsymbol{X})\boldsymbol{A} = (\boldsymbol{B} + \boldsymbol{\mathcal{E}}_B(\boldsymbol{X}))\boldsymbol{A} = \boldsymbol{I} + \boldsymbol{\mathcal{E}}_B(\boldsymbol{X})\boldsymbol{A} \tag{22}$$

where $\boldsymbol{\mathcal{E}}_B(\boldsymbol{X})$ is the estimation error in $\boldsymbol{B}$. In particular, when $\boldsymbol{A} = \boldsymbol{I}$ (and $\boldsymbol{X} = \boldsymbol{S}$) we have $\boldsymbol{T}(\boldsymbol{S}) = \boldsymbol{I} + \boldsymbol{\mathcal{E}}_B(\boldsymbol{S})$, where the variances of the elements of $\boldsymbol{\mathcal{E}}_B(\boldsymbol{S})$ are bounded by the respective CRLB, as expressed in (20) above. Recalling the invariance property of the iCRLB, and combining the estimation bound (21) with the ISR expression (2), we get (substituting $E[\boldsymbol{s}_k^T\boldsymbol{s}_k] = \mathrm{Tr}\{\boldsymbol{C}_k\}$)

$$\mathrm{ISR}_{k,\ell} \geq \frac{1}{N} \cdot \frac{\phi_{k,\ell}}{\phi_{k,\ell}\phi_{\ell,k} - 1} \cdot \frac{\mathrm{Tr}\{\boldsymbol{C}_\ell\}}{\mathrm{Tr}\{\boldsymbol{C}_k\}} \quad 1 \leq k \neq \ell \leq K \tag{23}$$

with $\phi_{k,\ell}$ defined in (19).

It is relatively straightforward (and reassuring) to show, that in the particular case of stationary sources, this bound coincides with asymptotic bounds developed previously under the stationarity assumption. These are, for example, the performance analysis of QML (with optimal filters) in [18], or the bound developed in [9] for the case of parametric (AR, MA, ARMA) Gaussian sources.

Some key properties of our iCRLB are summarized below.

- **Invariance w.r.t. $\boldsymbol{A}$**: As already mentioned, the iCRLB does not depend on the mixing matrix $\boldsymbol{A}$, but only on the sources' covariance matrices.
- **Invariance w.r.t. other sources**: The bound on $\mathrm{ISR}_{k,\ell}$ depends only on the covariance matrices of sources k and ℓ, and is unaffected by the other sources.
- **Invariance w.r.t. scale**: The ISR bound is invariant to any scaling of the sources. Note that this property is not shared by the bound on the variance of elements of $\boldsymbol{T}(\boldsymbol{X})$: If, for example, source k is amplified by a certain factor, the bounds on the variances of all of the resulting $T_{k,\ell}(\boldsymbol{X})$ $(k \neq \ell)$ would be reduced by the square of that factor; However, the bounds on all $\mathrm{ISR}_{k,\ell}$ would remain unchanged.
- **Non-identifiability condition**: If sources k and ℓ have similar covariance matrices (i.e., $\boldsymbol{C}_k$ is a scaled version of $\boldsymbol{C}_\ell$), then $\phi_{k,\ell} = 1/\phi_{\ell,k}$, implying (in (23)) an infinite bound on $\mathrm{ISR}_{k,\ell}$ and on $\mathrm{ISR}_{\ell,k}$—which in turn implies non-identifiability of the respective elements of $\boldsymbol{A}$. Recall that this bound was developed for Gaussian sources, and therefore this is a strict identifiability condition in the Gaussian case. In the case of non-Gaussian sources this identifiability condition can be easily shown to still be applicable to estimators based exclusively on SOS; However, when this condition is breached, mixtures of non-Gaussian sources may still be identifiable using HOS.
- **Resemblance to other bounds**: The general form of (21) is also shared by similar bounds developed for the case of sources with *i.i.d.* temporal structures (e.g., by Tichavský *et al.* in [22], or by Ollila *et al.* in [12]), with $\phi_{k,\ell}$ replaced by a quantity which depends only on the probability distribution function of the kth source.

We now turn to the derivation of the ML estimate.

B. The ML Estimate of $\boldsymbol{B}$

Recall from (4) that the $KN \times 1$ vector of concatenated mixture signals $\boldsymbol{x}$ is a zero-mean Gaussian vector, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C_x})$, with $\boldsymbol{C_x}$ given in (8). It therefore follows that the log-likelihood $\mathcal{L}(\boldsymbol{x};\boldsymbol{B})$ is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{x};\boldsymbol{B}) &= c - \frac{1}{2}\left(\log|\det \boldsymbol{C_x}| + \boldsymbol{x}^T\boldsymbol{C_x}^{-1}\boldsymbol{x}\right)\\ &= c' + N\log|\det \boldsymbol{B}|\\ &\quad - \frac{1}{2}\boldsymbol{x}^T\left(\boldsymbol{B}^T\otimes\boldsymbol{I}_N\right)\boldsymbol{C}_s^{-1}\left(\boldsymbol{B}\otimes\boldsymbol{I}_N\right)\boldsymbol{x} \qquad (24)\end{aligned}$$

where c and c' are constants which do not depend on $\boldsymbol{B}$. Differentiating w.r.t. $B_{k,\ell}$ we get (using the relation $\partial/\partial B_{k,\ell}\log|\det\boldsymbol{B}| = A_{\ell,k}$):

$$\begin{aligned}\frac{\partial\mathcal{L}(\boldsymbol{x};\boldsymbol{B})}{\partial B_{k,\ell}} &= NA_{\ell,k} - \frac{1}{2}\boldsymbol{x}^T\left[\left(\boldsymbol{E}_{\ell,k}\otimes\boldsymbol{I}_N\right)\boldsymbol{C_s}^{-1}\left(\boldsymbol{B}\otimes\boldsymbol{I}_N\right)\right.\\ &\qquad\left. + \left(\boldsymbol{B}^T\otimes\boldsymbol{I}_N\right)\boldsymbol{C_s}^{-1}\left(\boldsymbol{E}_{k,\ell}\otimes\boldsymbol{I}_N\right)\right]\boldsymbol{x}\\ &= NA_{\ell,k} - \frac{1}{2}\boldsymbol{x}^T\left[\left(\boldsymbol{E}_{\ell,k}\otimes\boldsymbol{C}_k^{-1}\right)\left(\boldsymbol{B}\otimes\boldsymbol{I}_N\right)\right.\\ &\qquad\left. + \left(\boldsymbol{B}^T\otimes\boldsymbol{I}_N\right)\left(\boldsymbol{E}_{k,\ell}\otimes\boldsymbol{C}_k^{-1}\right)\right]\boldsymbol{x}\\ &= NA_{\ell,k} - \frac{1}{2}\boldsymbol{x}^T\left[\left(\boldsymbol{e}_\ell\boldsymbol{b}_k^T + \boldsymbol{b}_k\boldsymbol{e}_\ell^T\right)\otimes\boldsymbol{C}_k^{-1}\right]\boldsymbol{x} \qquad (25)\end{aligned}$$

where $\boldsymbol{b}_k^T$ denotes the kth row of $\boldsymbol{B}$, namely $\boldsymbol{B} = [\boldsymbol{b}_1\cdots\boldsymbol{b}_K]^T$. Exploiting the relations

$$\begin{aligned}\boldsymbol{x}^T\left[\left(\boldsymbol{e}_\ell\boldsymbol{b}_k^T\right)\otimes\boldsymbol{C}_k^{-1}\right]\boldsymbol{x} &= \sum_{m=1}^{K} B_{k,m}\cdot\left(\boldsymbol{x}_\ell^T\boldsymbol{C}_k^{-1}\boldsymbol{x}_m\right)\\ \boldsymbol{x}^T\left[\left(\boldsymbol{b}_k\boldsymbol{e}_\ell^T\right)\otimes\boldsymbol{C}_k^{-1}\right]\boldsymbol{x} &= \sum_{m=1}^{K} B_{k,m}\cdot\left(\boldsymbol{x}_m^T\boldsymbol{C}_k^{-1}\boldsymbol{x}_\ell\right) \qquad (26)\end{aligned}$$

as well as the symmetry of $\boldsymbol{C}_k$, we end up with the likelihood equations (in which $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ denote the ML estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$ (respectively), naturally satisfying $\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}^{-1}$) of the form

$$\begin{aligned}\left.\frac{\partial\mathcal{L}(\boldsymbol{x};\boldsymbol{B})}{\partial B_{k,\ell}}\right|_{\hat{\boldsymbol{A}},\hat{\boldsymbol{B}}} &= N\hat{A}_{\ell,k} - \sum_{m=1}^{K}\hat{B}_{k,m}\cdot\left(\boldsymbol{x}_m^T\boldsymbol{C}_k^{-1}\boldsymbol{x}_\ell\right)\\ &= 0 \quad \forall k,\ \ell\in[1,K]. \qquad (27)\end{aligned}$$

Defining the $K\times K$ matrices

$$\hat{\boldsymbol{R}}^{(k)} \triangleq \frac{1}{N}\boldsymbol{X}\boldsymbol{C}_k^{-1}\boldsymbol{X}^T \quad k\in[1,K], \qquad (28)$$

we observe that the likelihood-equations (27) take the form

$$\hat{\boldsymbol{b}}_k^T\hat{\boldsymbol{R}}_{:,\ell}^{(k)} = \hat{A}_{\ell,k} \quad \forall k,\ell\in[1,K] \qquad (29)$$

where $\hat{\boldsymbol{b}}_k^T$ is the kth row of $\hat{\boldsymbol{B}}$ and where $\hat{\boldsymbol{R}}_{:,\ell}^{(k)}$ denotes the ℓth column of $\hat{\boldsymbol{R}}^{(k)}$. This implies

$$\hat{\boldsymbol{b}}_k^T\hat{\boldsymbol{R}}^{(k)} = [\hat{A}_{1,k}\ \hat{A}_{2,k}\cdots\hat{A}_{K,k}] = \boldsymbol{a}_k^T \qquad (30)$$

where $\hat{\boldsymbol{a}}_k$ denotes the kth column of $\hat{\boldsymbol{A}}$, namely $\hat{\boldsymbol{A}} = [\hat{\boldsymbol{a}}_1\cdots\hat{\boldsymbol{a}}_K]$. Exploiting the relation $\hat{\boldsymbol{B}}\hat{\boldsymbol{A}} = \boldsymbol{I}$, which implies that $\hat{\boldsymbol{b}}_m^T\hat{\boldsymbol{a}}_k = \delta_{m,k}$, we obtain the condition $\hat{\boldsymbol{b}}_m^T\hat{\boldsymbol{R}}^{(k)}\hat{\boldsymbol{b}}_k = \delta_{m,k}$, which can also be cast as

$$\hat{\boldsymbol{B}}\hat{\boldsymbol{R}}^{(k)}\hat{\boldsymbol{B}}^T\boldsymbol{e}_k = \boldsymbol{e}_k \Leftrightarrow \boldsymbol{e}_k^T\hat{\boldsymbol{B}}\hat{\boldsymbol{R}}^{(k)}\hat{\boldsymbol{B}}^T\boldsymbol{e}_\ell = \delta_{k,\ell} \quad k=1,\ldots,K. \qquad (31)$$

Note that this condition has an interesting interpretation, which is closely related to the concept of (non-orthogonal) approximate joint diagonalization (AJD), a well-studied subject in recent years (e.g., in [2], [15], [19], [23], [25], and [27], to name just a few). In the AJD framework, a set of P symmetric, $K\times K$ "target-matrices" $\boldsymbol{R}_1,\ldots,\boldsymbol{R}_P$ is given, and the problem is to find a matrix $\hat{\boldsymbol{B}}$ such that the transformed matrices $\hat{\boldsymbol{B}}\boldsymbol{R}_p\hat{\boldsymbol{B}}^T$ (for $p = 1,\ldots,P$) are "as diagonal as possible." Different criteria have been proposed for measuring the diagonality of the transformed set, leading to various optimization algorithms for finding $\hat{\boldsymbol{B}}$. In our case the set of "target matrices" consists of exactly $P = K$ matrices $\hat{\boldsymbol{R}}^{(1)},\ldots,\hat{\boldsymbol{R}}^{(K)}$, and the "diagonality criterion" reflects a hybrid "exact-approximate" diagonality requirement: The kth transformed matrix $\tilde{\boldsymbol{R}}^{(k)} \triangleq \hat{\boldsymbol{B}}\hat{\boldsymbol{R}}^{(k)}\hat{\boldsymbol{B}}^T$ is required to be exactly diagonal w.r.t. its kth column and row, meaning that this column and this row (only) must be all-zeros (exactly) except for the (k,k)th (diagonal) element, which should be 1; All other values in $\tilde{\boldsymbol{R}}^{(k)}$ are irrelevant. This should be satisfied for all the K matrices, namely for $k=1,\ldots,K$.

Such a "hybrid exact-approximate diagonalization" problem (which we term "HEAD") has already been encountered by (at least) Pham and Garat in [18] and Dégerine and Zaïdi in [8], in the context of (Q)ML separation of stationary sources (which is merely a particular case of our more general framework). The HEAD problem is discussed in detail in [28], where several iterative solution algorithms are outlined. For completeness of the exposition, we mention here the simple, yet rather effective algorithm by Dégerine and Zaïdi ([8]), which consists of alternating updates of the rows of $\hat{\boldsymbol{B}}$ as follows. Given some initial guess of $\hat{\boldsymbol{B}}$, repeat the following "sweep" (for $k=1,\ldots,K$):

$$\begin{aligned}\hat{\boldsymbol{b}}_k' &\leftarrow \hat{\boldsymbol{b}}_k - \hat{\boldsymbol{B}}_{(k)}^T\left(\hat{\boldsymbol{B}}_{(k)}\hat{\boldsymbol{R}}^{(k)}\hat{\boldsymbol{B}}_{(k)}^T\right)^{-1}\hat{\boldsymbol{B}}_{(k)}\hat{\boldsymbol{R}}^{(k)}\hat{\boldsymbol{b}}_k\\ \hat{\boldsymbol{b}}_k &\leftarrow \frac{\hat{\boldsymbol{b}}_k'}{\sqrt{\hat{\boldsymbol{b}}_k'^T\hat{\boldsymbol{R}}^{(k)}\hat{\boldsymbol{b}}_k'}} \qquad (32)\end{aligned}$$

where $\hat{\boldsymbol{B}}_{(k)}$ denotes the matrix $\hat{\boldsymbol{B}}$ without its kth row (a $(K-1)\times K$ matrix). When such "sweeps" are repeated iteratively, convergence to a solution of (31) is attained within several iterations (see [18] or [28] for more details).

To summarize, the ML estimate of $\boldsymbol{B}$ in the semi-blind case, in which the K sources' covariance matrices $\boldsymbol{C}_k$ are known, is obtained as follows:

1) for $k=1,\ldots,K$, compute the $K\times K$ "target-matrices" $\hat{\boldsymbol{R}}^{(k)} = 1/N\boldsymbol{X}\boldsymbol{C}_k^{-1}\boldsymbol{X}^T$;
2) find the matrix $\hat{\boldsymbol{B}}$, such that the K transformed matrices $\tilde{\boldsymbol{R}}^{(k)} = \hat{\boldsymbol{B}}\hat{\boldsymbol{R}}^{(k)}\hat{\boldsymbol{B}}^T$ each satisfies $\tilde{\boldsymbol{R}}^{(k)}\boldsymbol{e}_k = \boldsymbol{e}_k$ (namely its kth column is $\boldsymbol{e}_k$), e.g., using (32).

IV. The Fully-Blind Case

So far, we assumed that the sources' covariance matrices $\boldsymbol{C}_k$ are all known. Naturally, in practice this is rarely a realistic assumption. In this section we consider the "fully blind" scenario, in which these matrices are not known in advance. We shall assume a general model, reflecting the full range of possible blindness, as follows. Assume that the sources' covariance matrix $\boldsymbol{C_s}$ depends on some unknown parameters vector $\boldsymbol{\theta}$. Typically (but not necessarily), each covariance matrix $\boldsymbol{C}_k$ would be known only up to its respective parameters vector $\boldsymbol{\theta}_k$, such that $\boldsymbol{\theta} \triangleq [\boldsymbol{\theta}_1^T \cdots \boldsymbol{\theta}_K^T]^T$ is the concatenation of all these smaller vectors. The dimension of each $\boldsymbol{\theta}_k$ may range anywhere between 1 (e.g., when $\boldsymbol{C}_k$ is fully known up to scale, in which case $\boldsymbol{\theta}_k$ simply contains the unknown scale factor) and $N(N+1)/2$, in which case $\boldsymbol{\theta}_k$ may contain all the free unknown elements of the (symmetric) matrix $\boldsymbol{C}_k$.

Define a vector containing the entire set of unknown (mixing/demixing and covariances) parameters $\boldsymbol{\xi} \triangleq [\text{vec}^T(\boldsymbol{B})\ \boldsymbol{\theta}^T]^T$. We shall now show that under some scaling assumption (to be specified shortly), the FIM $\boldsymbol{J}_{\boldsymbol{\xi}}$ (w.r.t. $\boldsymbol{\xi}$) is block-diagonal, with different blocks accounting for $\text{vec}(\boldsymbol{B})$ and for $\boldsymbol{\theta}$. It is important to note here, that although the iCRLB is invariant in $\boldsymbol{B}$ (also in the fully blind case), the FIM $\boldsymbol{J}_{\boldsymbol{\xi}}(\boldsymbol{B})$ is certainly not invariant in $\boldsymbol{B}$, and therefore it would not be sufficient to establish block-diagonality of $\boldsymbol{J}_{\boldsymbol{\xi}}(\boldsymbol{B})$ for $\boldsymbol{B} = \boldsymbol{I}$ only. We need to show that $\boldsymbol{J}_{\boldsymbol{\xi}}(\boldsymbol{B})$ is block-diagonal for all (nonsingular) $\boldsymbol{B}$. Let θ denote an arbitrary element of $\boldsymbol{\theta}$. Recall that the general expression for the cross-blocks elements of the FIM is given by

$$\begin{aligned} J_{\theta,B_{k,\ell}}(\boldsymbol{B}) &= \frac{1}{2}\text{Tr}\left\{\frac{\partial \boldsymbol{C}_{\boldsymbol{x}}}{\partial\theta}\cdot\boldsymbol{C}_{\boldsymbol{x}}^{-1}\cdot\frac{\partial \boldsymbol{C}_{\boldsymbol{x}}}{\partial B_{k,\ell}}\cdot\boldsymbol{C}_{\boldsymbol{x}}^{-1}\right\} \\ &= -\frac{1}{2}\text{Tr}\left\{\frac{\partial \boldsymbol{C}_{\boldsymbol{x}}}{\partial\theta}\cdot\frac{\partial \boldsymbol{C}_{\boldsymbol{x}}^{-1}}{\partial B_{k,\ell}}\right\} \\ &= -\frac{1}{2}\text{Tr}\left\{\frac{\partial \boldsymbol{C}_{\boldsymbol{x}}^{-1}}{\partial B_{k,\ell}}\cdot\boldsymbol{C}'_{\boldsymbol{x}}\right\} \end{aligned} \tag{33}$$

where we have used the identity (11), and where

$$\boldsymbol{C}'_{\boldsymbol{x}} \triangleq \frac{\partial}{\partial\theta}\boldsymbol{C}_{\boldsymbol{x}} = \left(\boldsymbol{B}^{-1}\otimes\boldsymbol{I}_N\right)\boldsymbol{C}'_{\boldsymbol{s}}\left(\boldsymbol{B}^{-T}\otimes\boldsymbol{I}_N\right)^T \tag{34}$$

($\boldsymbol{C}'_{\boldsymbol{s}}$ being used as shorthand for $\partial/\partial\theta\boldsymbol{C}_{\boldsymbol{s}}$).

Exploiting the structural similarity between $\boldsymbol{C}_{\boldsymbol{x}}$ and $\boldsymbol{C}'_{\boldsymbol{x}}$, we can substitute $\boldsymbol{C}_{\boldsymbol{s}}$ with $\boldsymbol{C}'_{\boldsymbol{s}}$ in (13) in order to obtain

$$\frac{\partial \boldsymbol{C}_{\boldsymbol{x}}^{-1}}{\partial B_{k,\ell}}\cdot\boldsymbol{C}'_{\boldsymbol{x}} = \left([\boldsymbol{E}_{\ell,k}+\boldsymbol{B}^T\boldsymbol{E}_{k,\ell}\boldsymbol{B}^{-1}]\otimes\boldsymbol{C}_k^{-1}\right)\boldsymbol{C}'_s\left(\boldsymbol{B}^{-T}\otimes\boldsymbol{I}_N\right). \tag{35}$$

Therefore,

$$\begin{aligned} &\text{Tr}\left\{\frac{\partial \boldsymbol{C}_{\boldsymbol{x}}^{-1}}{\partial B_{k,\ell}}\cdot\boldsymbol{C}'_{\boldsymbol{x}}\right\} \\ &\quad= \text{Tr}\left\{\left(\boldsymbol{B}^{-T}\otimes\boldsymbol{I}_N\right)\left([\boldsymbol{E}_{\ell,k}+\boldsymbol{B}^T\boldsymbol{E}_{k,\ell}\boldsymbol{B}^{-1}]\otimes\boldsymbol{C}_k^{-1}\right)\boldsymbol{C}'_s\right\} \\ &\quad= \text{Tr}\left\{\left([\boldsymbol{B}^{-T}\boldsymbol{E}_{\ell,k}+\boldsymbol{E}_{k,\ell}\boldsymbol{B}^{-1}]\otimes\boldsymbol{C}_k^{-1}\right)\boldsymbol{C}'_s\right\} \\ &\quad= \text{Tr}\left\{\left([\boldsymbol{E}_{k,\ell}\boldsymbol{B}^{-1}]\otimes\boldsymbol{C}_k^{-1}\right)\boldsymbol{C}'_s\right\} \\ &\qquad\quad+\text{Tr}\left\{\boldsymbol{C}'_s\left([\boldsymbol{B}^{-T}\boldsymbol{E}_{\ell,k}]\otimes\boldsymbol{C}_k^{-1}\right)\right\}. \end{aligned} \tag{36}$$

Due to the symmetry of $\boldsymbol{C}_k$ and of $\boldsymbol{C}'_s$, these two traces are equal, so

$$\begin{aligned} J_{\theta,B_{k,\ell}}(\boldsymbol{B}) &= -\frac{1}{2}\text{Tr}\left\{\frac{\partial \boldsymbol{C}_{\boldsymbol{x}}^{-1}}{\partial B_{k,\ell}}\cdot\boldsymbol{C}'_{\boldsymbol{x}}\right\} \\ &= \text{Tr}\left\{\left([\boldsymbol{E}_{k,\ell}\boldsymbol{A}]\otimes\boldsymbol{C}_k^{-1}\right)\boldsymbol{C}'_s\right\}. \end{aligned} \tag{37}$$

Like $\boldsymbol{C}_s$, $\boldsymbol{C}'_s$ is block-diagonal, so only the respective diagonal blocks of $[\boldsymbol{E}_{k,\ell}\boldsymbol{A}]\otimes\boldsymbol{C}_k^{-1}$ are relevant for the trace. These are determined by the diagonal terms of $\boldsymbol{E}_{k,\ell}\boldsymbol{A} = \boldsymbol{e}_k\boldsymbol{e}_\ell^T\boldsymbol{A}$. The only nonzero element along the principal diagonal of this matrix can be its (k,k)th element, which equals $A_{\ell,k}$. The respective (k,k)th block of $\boldsymbol{C}'_{\boldsymbol{s}}$ is $\boldsymbol{C}'_k$, and therefore

$$J_{\theta,B_{k,\ell}}(\boldsymbol{B}) = -A_{\ell,k}\text{Tr}\{\boldsymbol{C}_k^{-1}\boldsymbol{C}'_k\}. \tag{38}$$

Noting that $\text{Tr}\{\boldsymbol{C}_k^{-1}\boldsymbol{C}'_k\}$ is simply the derivative of $\log\det\boldsymbol{C}_\ell$ w.r.t. θ, we conclude that the FIM $\boldsymbol{J}_{\boldsymbol{\xi}}$ is block-diagonal if for each $k=1,\ldots,K$, the *determinant* $\det\boldsymbol{C}_k$ does not depend on any of the unknown parameters in $\boldsymbol{\theta}$ (but, of course, $\boldsymbol{C}_k$ itself may certainly depend on some or on all of the parameters in $\boldsymbol{\theta}$).

A common example of a case where $\boldsymbol{C}_k$ depends on a parameters vector $\boldsymbol{\theta}$ but its determinant does not is the following (for convenience, we drop the index k in this example). Consider a source signal $\boldsymbol{s}$ whose elements $s[0],\ldots,S[N-1]$ satisfy the difference equations

$$s[n] = -\sum_{q=1}^{Q}a_q[n]s[n-q]+v[n] + \sum_{p=1}^{P}b_p[n]v[n-p] \quad n=0,\ldots,N-1 \tag{39}$$

where the "driving noise" $v[n]$ is a zero-mean white process with fixed (known) variance σ_v^2 (and employing the conventions $s[n]=0$, $v[n]=0$ for $n<0$). The vector of unknown relevant[3] parameters consists, in the most general case, of $\boldsymbol{\theta} = [\boldsymbol{a}^T\boldsymbol{b}^T]^T$ with (40),

$$\begin{aligned} \boldsymbol{a} &= [a_1[1]\cdots a_1[N-1]\ a_2[2]\cdots a_2[N-1] \\ &\qquad \cdots a_Q[Q]\cdots a_Q[N-1]]^T \\ \boldsymbol{b} &= [b_1[1]\cdots b_1[N-1]\ b_2[2]\cdots b_2[N-1] \\ &\qquad \cdots b_P[P]\cdots b_P[N-1]]^T \end{aligned} \tag{40}$$

with Q and P denoting, respectively, the AR and MA orders of the process. These difference equations define a nonstationary ARMA process with possibly time-varying AR and MA coefficients (the unknown parameters). To see that $\det\boldsymbol{C}$ does not depend on $\boldsymbol{\theta}$ in this case, note that the difference (39) can be expressed in matrix-vector form as

$$\boldsymbol{G}(\boldsymbol{a})\boldsymbol{s} = \boldsymbol{H}(\boldsymbol{b})\boldsymbol{v} \tag{41}$$

where $\boldsymbol{v} = [v[0]\cdots v[N-1]]^T$ is the "driving-noise" vector and $\boldsymbol{G}(\boldsymbol{a})$ and $\boldsymbol{H}(\boldsymbol{b})$ are lower-triangular banded matrices with 1-s along their principal diagonals. We thus have $\boldsymbol{s} = \boldsymbol{G}^{-1}(\boldsymbol{a})\boldsymbol{H}(\boldsymbol{b})\boldsymbol{v}$, so

$$\boldsymbol{C}(\boldsymbol{\theta}) = E[\boldsymbol{s}\boldsymbol{s}^T] = \boldsymbol{G}(\boldsymbol{a})^{-1}\boldsymbol{H}(\boldsymbol{b})\boldsymbol{C}_{\boldsymbol{v}}\boldsymbol{H}^T(\boldsymbol{b})\boldsymbol{G}^{-T}(\boldsymbol{a}) \tag{42}$$

where $\boldsymbol{C}_{\boldsymbol{v}} = E[\boldsymbol{v}\boldsymbol{v}^T] = \sigma_v^2\boldsymbol{I}_N$. Since $\boldsymbol{G}(\boldsymbol{a})$ and $\boldsymbol{H}(\boldsymbol{b})$ are lower-triangular with a fixed all-ones diagonal, their determinants (as well as the determinants of their inverses) are 1-s, and therefore $\det\boldsymbol{C}(\boldsymbol{\theta}) = \sigma_v^{2N}$ and does not depend on $\boldsymbol{\theta}$ (note that such independence also holds with any other driving-noise covariance $\boldsymbol{C}_v$ whose determinant is fixed, but the case $\boldsymbol{C}_v = \sigma_v^2\boldsymbol{I}_N$ is considered more "standard").

Naturally, one widely familiar particular case within this framework is the case of constant regression coefficients, which occurs when $a_p[n] = a_p$ and $b_q[n] = b_q$ for all p, q, n. Note,

[3]Note that $a_p[n]$ and $b_q[n]$ for $p > n$ and for $q > n$, respectively, are irrelevant, since they are always multiplied by zeros.

however, that strictly speaking, this does *not* lead to a stationary signal $\boldsymbol{s}$, since the difference (39) entails the effect of zero initial conditions (in other words, $\boldsymbol{C}(\boldsymbol{\theta})$ in (42) is *not* a Toeplitz matrix in this case). Nevertheless, asymptotically the effect of initial conditions (or the deviation of $\boldsymbol{C}(\boldsymbol{\theta})$ from a Toeplitz structure) becomes negligible, and $\boldsymbol{s}$ can be considered as a segment taken from a stationary ARMA process. It is interesting to observe, that this is the reason why the CRLB expression obtained in [8] on estimation of the demixing matrix $\boldsymbol{B}$ for stationary Gaussian AR sources is only asymptotic: when the sources are truly stationary, the FIM is not block-diagonal, since the determinants of all $\boldsymbol{C}_k$-s depend on the respective unknown AR parameters. Only asymptotically (as $N \longrightarrow \infty$) this dependence becomes negligible, as the fully stationary model becomes essentially equivalent to our zero initial-conditions model above.

We therefore conclude that, since we have shown the FIM to be block-diagonal, the CRLB on estimation of $\boldsymbol{B}$ (or $\boldsymbol{A}$) when the covariance matrices $\boldsymbol{C}_k$ are *known* is the same as the CRLB obtained when these matrices are *unknown*, provided that their determinants are known. For the same to hold for the iCRLB, however, knowledge of these determinants is not required. To show this, we recall that the iCRLB (as opposed to the CRLB) is invariant in $\boldsymbol{A}$, and may therefore be calculated with any chosen (nonsingular) $\boldsymbol{A}$, with the result holding true for all other (nonsingular) $\boldsymbol{A}$-s.

Choosing $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{I}$ again, the only remaining (nonzero) off-block-diagonal terms of the FIM in (38) in this case are terms related to $J_{\theta, B_{k,k}}(\boldsymbol{I})$. Recalling the general block-diagonal structure (18) of the "$\boldsymbol{B}$-part" of $\boldsymbol{J}_{\boldsymbol{\xi}}(\boldsymbol{I})$, this means that when $\boldsymbol{A} = \boldsymbol{I}$, the only element of $\boldsymbol{B}$ whose estimation bound may be affected (increased) by lack of knowledge of $\det \boldsymbol{C}_k$ is (according to (38)) $B_{k,k}$.

Fortunately, still thanks to the block-diagonality of the "$\boldsymbol{B}$-part" of $\boldsymbol{J}_{\boldsymbol{\xi}}(\boldsymbol{I})$ in (18), this has no effect on the ISR bound (23). Note that this is a rather intuitive result, since uncertainty in $\det \boldsymbol{C}_k$ implies uncertainty in the scale of the kth source, which is well-known to be unresolvable in BSS on one hand, but to have no effect on the ISR on the other hand.

We therefore conclude that the iCRLB is indifferent to the knowledge of the sources' covariance matrices (or of any related parameters such as their scales or determinants). In other words, the iCRLB expression (23), which was obtained for the semi-blind scenario, is also valid in the fully blind scenario.

Of course, this does not mean in general that the same ISR can be attained in the semi-blind and fully blind cases, but it does have an important implication on the asymptotic performance in a multiple-experiments scenario: Indeed, assume that multiple independent experiments (or "multiple snapshots") $\boldsymbol{X}_m = \boldsymbol{A}\boldsymbol{S}_m$ are available (for $m = 1, \ldots, M$), where in each experiment the sources are redrawn from the same distribution (with the same $\boldsymbol{C}_k$-s). For example, such a situation can occur when conducting a sequence of repeated experiments, such that each experiment is synchronized to some external trigger, causing each source to obey the same temporal-covariance pattern (e.g., some energy rise-time and/or fall-time, some triggered frequency-drift, etc.) in each experiment.

Then, due to the asymptotic efficiency of the ML estimate, the block-diagonality of $\boldsymbol{J}_{\boldsymbol{\xi}}$ implies that asymptotically (as $M \longrightarrow \infty$) the same ISR which can be attained in the semi-blind scenario (when all $\boldsymbol{C}_k$ are known in advance) can also be attained in the fully blind scenario (when they are not known in advance), and both coincide with the iCRLB developed earlier.

Intuitively, the ability to match the semi-blind performance asymptotically in the fully blind scenario is based on the ability to effectively estimate the sources' covariance matrices from the observed mixtures over the multiple experiments. In particular, an iterative process can be proposed as follows.

1) Apply some "standard" separation algorithm (e.g., second-order blind identification (SOBI) [4]) to the concatenated observations $\tilde{\boldsymbol{X}} \triangleq [\boldsymbol{X}_1 \cdots \boldsymbol{X}_M]$, denoting the estimated separation matrix as $\hat{\boldsymbol{B}}$.
2) Extract the M estimated $K \times N$ source matrices, $\hat{\boldsymbol{S}}_m = \hat{\boldsymbol{B}}\boldsymbol{X}_m, \quad m = 1, \ldots, M$, and denote the $M \cdot K$ extracted $N \times 1$ signals as $\hat{\boldsymbol{s}}_{k,m}$ (namely $\hat{\boldsymbol{S}}_m = [\hat{\boldsymbol{s}}_{1,m} \cdots \hat{\boldsymbol{s}}_{K,m}]^T$).
3) Estimate the covariance matrices, e.g., $\hat{\boldsymbol{C}}_k = 1/M \sum_{m=1}^{M} \hat{\boldsymbol{s}}_{k,m}\hat{\boldsymbol{s}}_{k,m}^T, \quad k = 1, \ldots, K$.
4) Using $\hat{\boldsymbol{C}}_k$ as substitutes for the true $\boldsymbol{C}_k$, apply the ML estimation scheme described at the end of in Section III and obtain an updated estimate $\hat{\boldsymbol{B}}$ of the demixing matrix.
5) Go back to step 2 and repeat until convergence is attained.

While such a scheme is theoretically feasible, it would generally require a huge number of experiments $MN \gg N^2$, namely $M \gg N$, in order to obtain reliable estimates of the general covariance matrices. Fortunately, however, in many cases of interest, the sources may have a general, yet succinctly parameterized covariance structure, thereby significantly reducing the number M of required experiments, so as to match the (small) number of free parameters. For example, with TVAR [1] sources, a very small number of experiments $(M \ll N)$ can be sufficient for reliable estimation of the TVAR parameters (which determine the sources' covariances), attaining optimal performance (i.e., matching the iCRLB and the semi-blind performance). In fact, in some cases (e.g., stationary or cyclostationary sources), a single (sufficiently long) realization of each source signal is sufficient for obtaining a reliable estimate of its covariance. Therefore, the iterative scheme can be useful also in the single-experiment $(M = 1)$ scenario, using suitable estimation schemes for estimating the parameters of the covariance matrices from the extracted sources. We demonstrate both kinds of scenarios in simulation results in the next section.

We note in passing, that when the matrices $\boldsymbol{C}_k$ are not constrained to have a special structure, an alternative approach for obtaining the multiple-experiments (blind) ML estimate of $\boldsymbol{B}$ can be taken: In this case maximizing the likelihood is equivalent to minimizing the Kullback-Leibler divergence (KLD) between the empirical $(KN \times KN)$ covariance matrix[4] of the estimated sources $\hat{\boldsymbol{C}}_{\boldsymbol{S}} \triangleq 1/M \sum_{m=1}^{M} \hat{\boldsymbol{s}}_m \hat{\boldsymbol{s}}_m^T$ (where $\hat{\boldsymbol{s}}_m = [\hat{\boldsymbol{s}}_{1,m}^T \cdots \hat{\boldsymbol{s}}_{K,m}^T]^T = \mathrm{vec}(\hat{\boldsymbol{S}}_m^T)$) and its block-diagonal version (consisting only of its K diagonal blocks of size $N \times N$). This can be attained using the iterative joint block-diagonalization algorithm proposed by Pham (in a different context) in [16] (which, in our case, would be applied only to a single matrix $\hat{\boldsymbol{C}}_{\boldsymbol{S}}$).

V. Simulation Results

To demonstrate the theoretical results and the performance of the proposed estimates in their respective contexts, we present

[4]We loosely refer to the KLD between two zero-mean Gaussian distribution as the KLD between their covariance matrices.

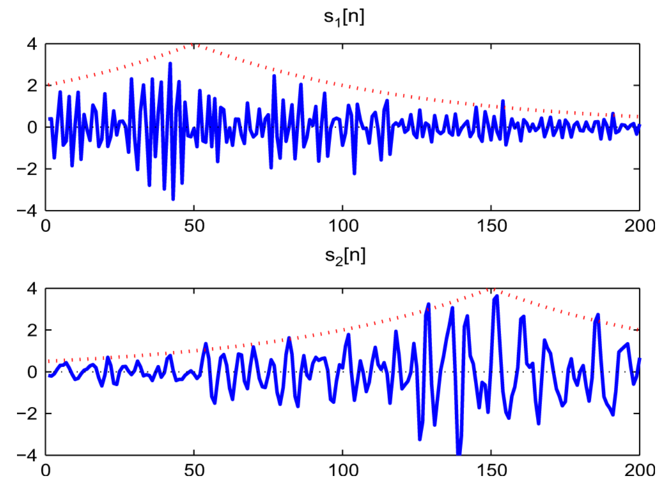

Fig. 1. Sample functions of the two source signals used in the first experiment, expressing both spectral and temporal localization diversities.

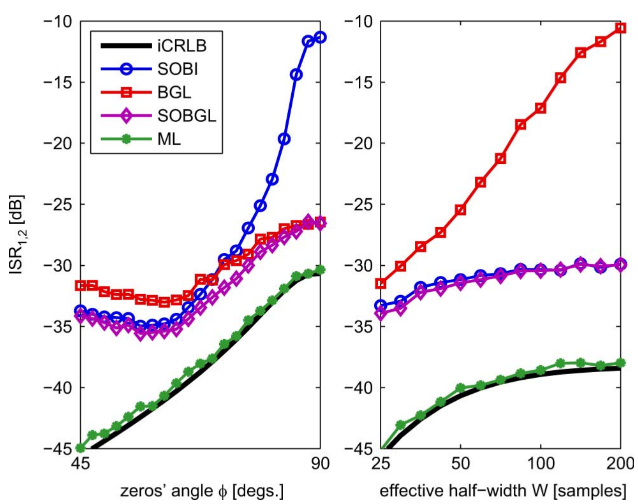

Fig. 2. $\mathrm{ISR}_{1,2}$ versus ϕ and W: iCRLB and the empirical performance of SOBI, BGL and SOBGL.

simulation results of three different experiments. The first experiment addresses a semi-blind scenario, whereas the other two address (nearly) fully blind scenarios: one with a "single experiment" ($M = 1$) scenario and the other with a "multiple experiments" scenario. In all three experiments, each empirical result represents an average of 1000 independent trials, with the mixing matrix drawn at random (with independent zero-mean unit-variance Gaussian elements) in each trial.

A. Experiment A: MA Sources With a Laplacian Envelope

In the first experiment we use two non-stationary source signals generated as follows. First, we generate two stationary Gaussian MA signals, one with four zeros at $z = 0.8, -0.9, 0.9e^{\pm j\phi}$ and the other with four zeros at $z = 0.9, -0.8, -0.9e^{\pm j\phi}$ (and their reciprocals), where ϕ is a parameter controlling the spectral diversity between these signals: as ϕ moves from 0 to $\pi/2$, the two spectra become more similar. Then a segment of $N = 200$ samples from each of these signals is multiplied by a Laplacian-shaped window, $w[n] = e^{-\log(0.5)/2W|n-N_c|}$, centered around $N_c = 50$ for the first signal and around $N_c = 150$ for the second signal, with variable width W (measured to the -3 dB points of the Laplacian window). Thus, W controls the temporal-location diversity of the signals: as W increases, the localization decreases, so that when $W \ll N$ the signals are practically distinct in time, whereas when $W \gg N$ there's nearly no temporal distinction.

An example of a sample function of the two signals for $\phi = \pi/4$ and $W = 25$ is depicted in Fig. 1 (with the Laplacian windows in dashed lines): The first signal has a relatively high-frequency content and its energy is concentrated more in the first half of the segment, whereas the second signal has a relatively low-frequency content, and its energy is concentrated more in the second half.

Evidently, the SOBI algorithm [4] can be expected to capture and exploit the spectral diversity of the underlying signals, but cannot exploit their different temporal concentrations. Conversely, the BGL algorithm [17] (with just two blocks, each of size $N/2 = 100$) can exploit the different temporal concentrations but cannot exploit the spectral diversity. Now, recall that SOBI is based on joint diagonalization of correlation matrices at different lags, whereas BGL is based on joint diagonalization of zero-lag correlations taken over different blocks. A natural extension which comes to mind is to combine SOBI and BGL into what we would call "SOBGL" for short, and would be based on joint diagonalization of all these correlation matrices together, hopefully being able to exploit both sources of diversity. Obviously, however, this is just an *ad-hoc* combination, which is far from being optimal—as we demonstrate in the simulation results on Fig. 2.

Fig. 2 shows the attained performance in terms of $\mathrm{ISR}_{1,2}$ (in decibels [dB]). The left-hand plot shows the ISR as ϕ varies from $\pi/4$ to $\pi/2$ with W fixed at 25, whereas the right-hand plot shows the ISR as W varies from 25 to 200 with ϕ fixed at $\pi/4$. We show the iCRLB, as well as the ISR attained by SOBI, BGL, SOBGL and our ML (implemented as prescribed for the semi-blind case in Section III-B above). On the left-hand plot we see that SOBI, which performs well with the smaller ϕ, deteriorates rapidly as ϕ approaches $\pi/2$ and the spectral diversity is nearly lost. BGL is relatively insensitive to ϕ, and SOBGL slightly outperforms the better of these two at each ϕ. However, all three are significantly worse than our ML estimates, which attains the iCRLB. On the right-hand plot we see that BGL deteriorates rapidly with the loss of localization diversity, whereas SOBI is relatively constant and SOBGL is slightly better. Again, our ML estimate significantly outperforms all three, coinciding with the iCRLB.

It is important to realize that, obviously, in this semi-blind scenario ML has a clear "unfair" advantage over SOBI, BGL and SOBGL, which, unlike ML, cannot exploit the prior knowledge of the sources' covariance matrices. Nevertheless, it is our purpose in this work to show just how such information can be best exploited when available. Moreover, in the fully blind experiments which follow, this "unfair" prior knowledge is eliminated.

B. Experiment B: AR Sources With a Cyclostationary Driving Noise

In the second experiment, we use four AR sources of order $P = 2$, generated using driving-noise sequences whose power-profiles change periodically. Therefore, the resulting sources are cyclostationary. More specifically, each source satisfies the difference equation

$$s_k[n] = -a_1^{(k)} s_k[n-1] - a_2^{(k)} s_k[n-2] + \tilde{v}_k[n],$$

TABLE I
SOURCES' PARAMETERS USED IN (43). NOTE THAT SINCE $A_3 = A_4 = 0$, ϕ_3, ϕ_4, T_3 AND T_4 ARE MEANINGLESS

	s_1	s_2	s_3	s_4
A_k	0.5	0.6	0	0
ϕ_k [deg.]	145^o	55^o	-	-
T_k	50	70	-	-
poles' magnitudes	0.75	0.7	0.7	0.75
poles' phases	$\pm j\frac{95}{180}\pi$	$\pm j\frac{85}{180}\pi$	$\pm j\frac{95}{180}\pi$	$\pm j\frac{85}{180}\pi$

with

$$\tilde{v}_k[n] \triangleq \sqrt{1 + A_k \cos\left(2\pi \cdot \frac{n}{T_k} + \phi_k\right)} \cdot v_k[n] \tag{43}$$

where $\{v_k[n]\}_{k=1}^4$ are i.i.d. zero-mean unit-variance Gaussian sequences, and where the coefficients $\{A_k, \phi_k, T_k\}_{k=1}^4$, as well as the poles of the polynomials $\{1+a_1^{(k)}z^{-1}+a_2^{(k)}z^{-2}\}_{k=1}^4$ are specified in Table I. It is evident that since $A_3 = A_4 = 0$, the third and fourth sources are stationary AR sources. In addition, the spectral diversity of all sources is rather weak (the poles are rather close), so good separation, e.g., between sources 1 and 3 and between sources 2 and 4, has to rely on the difference in the driving-noise envelopes.

We assume that the general structures of the sources are known in advance, but only up to the unknown parameters of Table I (this is why we term this a "nearly" fully blind scenario). The blind separation therefore proceeds as follows. First, we apply initial separation using SOBI with three correlation matrices (at lags 0, 1, 2). We then repeat the following for three iterations:

1) For each separated (estimated) source $\hat{s}_k[n]$ ($k = 1, 2, 3, 4$):
   - estimate the AR parameters $a_1^{(k)}$, $a_2^{(k)}$ using Yule–Walker equations (e.g., [11]) applied to ordinary correlation estimates of $\hat{s}_k[n]$;
   - using the estimated AR parameters, obtain the estimated driving-noise sequence via respective FIR filtering of $\hat{s}_k[n]$; denote this sequences as $\hat{\tilde{v}}_k[n]$;
   - using the discrete-time Fourier transform (DTFT) of the squared $\hat{\tilde{v}}_k^2[n]$, obtain an estimate $\hat{T}_k$ of the period, as the reciprocal of the (non-DC) highest peak-location.
   - using least-squares (LS) fit of $\hat{\tilde{v}}_k^2[n]$, estimate the amplitude A_k and phase ϕ_k (this can be attained using linear LS fit of a constant + sine and cosine sequences with the estimated periods $\hat{T}_k$).
2) Using the estimated parameters, which in turn yield the estimated sources' covariance matrices, obtain the (blind) ML estimates of the sources.

(note that estimation of T_k, A_k and ϕ_k was applied to all four sources, including the stationary sources (3 and 4), since the information that $A_3 = A_4 = 0$ is unknown to the estimator).

Fortunately, in our case the procedure does not require explicit computation and inversion of the estimated sources' covariance matrices (which are all $N \times N$), since the AR model already provides direct access to the inverse covariance, and the driving-noise envelope can also be easily inverted. More specifically, let $\hat{a}_1^{(k)}$, $\hat{a}_2^{(k)}$, $\hat{T}_k$, $\hat{A}_k$ and $\hat{\phi}_k$ denote the estimated parameters, and let $\hat{d}^2[n] \triangleq 1 + \hat{A}_k \cos(2\pi n/\hat{T}_k + \hat{\phi}_k)$ denote the (squared) estimated envelope of the driving-noise. Defining $\hat{\boldsymbol{A}}_k$ as the $N \times N$ Toeplitz matrix with 1-s along its main diagonal, $\hat{a}_1^{(k)}$ along its first sub- diagonal and $\hat{a}_1^{(k)}$ along its second sub-diagonal, and denoting by $\hat{\boldsymbol{D}}_k$ the $N \times N$ diagonal matrix with $\hat{d}_k[n]$ along its diagonal, the implied estimate of the covariance matrix is given by

$$\hat{\boldsymbol{C}}_k = \hat{\boldsymbol{A}}_k^{-1} \hat{\boldsymbol{D}}_k^2 \hat{\boldsymbol{A}}_k^{-T} \tag{44}$$

so the (blind) ML correlation matrices, given by

$$\hat{\boldsymbol{R}}^{(k)} = \frac{1}{N}\boldsymbol{X}\hat{\boldsymbol{C}}_k^{-1}\boldsymbol{X}^T = \frac{1}{N}(\boldsymbol{X}\hat{\boldsymbol{A}}_k^T\hat{\boldsymbol{D}}_k^{-1})(\boldsymbol{X}\hat{\boldsymbol{A}}_k^T\hat{\boldsymbol{D}}_k^{-1})^T \tag{45}$$

can be easily obtained as $\hat{\boldsymbol{R}}^{(k)} = 1/N\boldsymbol{W}^{(k)}\boldsymbol{W}^{(k)T}$, where each row of $\boldsymbol{W}^{(k)} \triangleq \boldsymbol{X}\hat{\boldsymbol{A}}_k^T\hat{\boldsymbol{D}}_k^{-1}$ is obtained by first filtering the respective row of $\boldsymbol{X}$ with the FIR filter whose taps are $[1\ \hat{a}_1^{(k)}\ \hat{a}_2^{(k)}]$ and then dividing the result (elementwise) by the sequence $\hat{d}_k[n]$.

This enables computation of the matrices $\hat{\boldsymbol{R}}^{(k)}$ in $\mathcal{O}(N)$, rather than $\mathcal{O}(N^3)$ operations, saving the need for explicit computation and inversion of the sources' estimated covariance matrices $\hat{\boldsymbol{C}}_k$.

In Fig. 3, we show the obtained $\text{ISR}_{k,\ell}$ values for all $k \neq \ell \in \{1, \ldots, 4\}$ versus the observation length N. In addition to the obtained blind-ML (BML) results, we also show the following for comparison:

- the respective iCRLBs;
- the (semi-blind) ML separation results, obtained assuming known covariance matrices of the sources, namely using the true parameters of Table I;
- the WASOBI [23] separation results (using the code in [20]).

WASOBI (an asymptotically optimally-weighted version of SOBI) is supposed to attain the iCRLB asymptotically for stationary AR sources, and indeed it can be seen to coincide (asymptotically) with the iCRLB for $\text{ISR}_{3,4}$ and for $\text{ISR}_{4,3}$, since sources 3 and 4 are stationary AR sources. However, for all other couples, which involve at least one nonstationary source, WASOBI is seen to be suboptimal, being unable to exploit the additional nonstationarity diversity. In particular, the WASOBI performance is severely suboptimal in resolving source 1 and 3, as well are 2 and 4, since these couples share very similar poles and differ mainly by their driving-noise envelopes. The ML separation is seen to already coincide with the respective iCRLB for relatively low values of N (typically less than 500), whereas the BML separation obviously needs more data to be able to obtain useful estimates of the covariance matrices, and therefore only coincides with the iCRLB for N larger than, say, 2000.

C. Experiment C: TVAR Sources

In the last experiment we use five time-varying AR sources of order $P = 4$, generated as follows:

$$s_k[n] = -\sum_{p=1}^{4} a_p^{(k)}[n]s_k[n-p] + v_k[n] \tag{46}$$

where the driving-noise sequences $v_k[n]$ are mutually independent, white zero-mean unit-variance Gaussian processes, and where the TVAR coefficients reflect a linear drift of the poles from an initial state (at $n = 0$) to a final state at $n = N - 1 =$

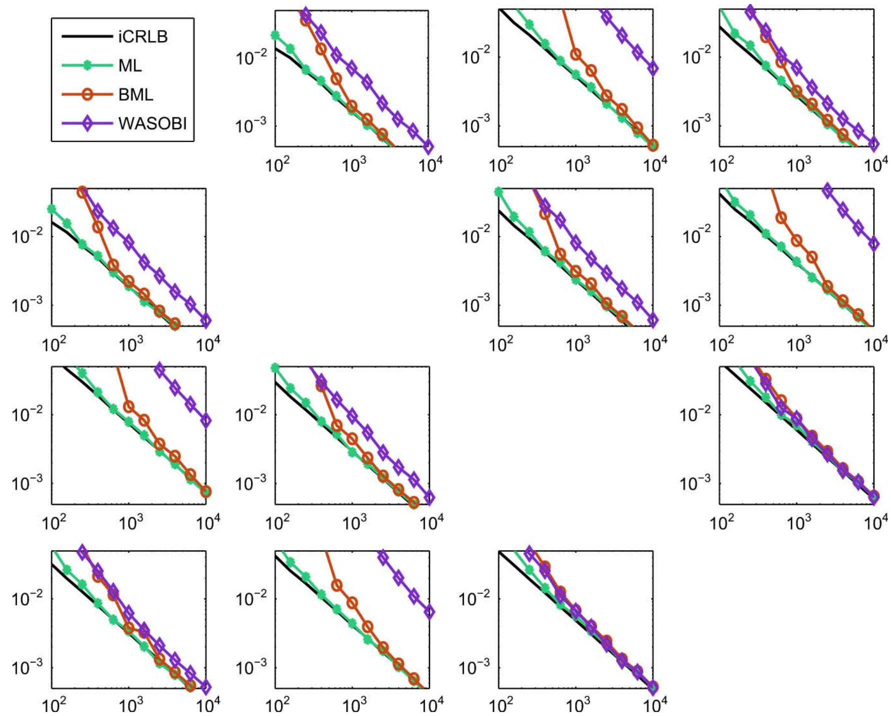

Fig. 3. $\mathrm{ISR}_{k\neq\ell}$ (for $k,\ell\in\{1,2,3,4\}$, k is the row-index and ℓ is the column-index) versus N: iCRLB, ML, BML and WASOBI.

TABLE II
INITIAL AND FINAL VALUES OF THE FOUR POLES FOR THE FIVE SOURCES. NOTE THAT THE FIFTH SOURCE IS STATIONARY

	s_1	s_2	s_3	s_4	s_5
$\varrho_1^{(k)}[0], \varrho_1^{(k)}[N]$	$0.8, 0.7$	$0.8, 0.7$	$0.7, 0.8$	$0.7, 0.8$	$0.7, 0.7$
$\varrho_2^{(k)}[0], \varrho_2^{(k)}[N]$	$-0.7, -0.8$	$-0.7, -0.8$	$-0.8, -0.7$	$-0.8, -0.7$	$-0.9, -0.9$
$\varrho_3^{(k)}[0], \varrho_3^{(k)}[N]$	$0.7e^{j\frac{\pi}{4}}, -0.9$	$0.9e^{j\frac{\pi}{4}}, 0.7j$	$0.7e^{j\frac{\pi}{4}}, 0.9$	$0.9e^{j\frac{\pi}{4}}, -0.7j$	$0.8e^{j\frac{\pi}{4}}, 0.8e^{j\frac{pi}{4}}$
$\varrho_4^{(k)}[0], \varrho_4^{(k)}[N]$	$0.7e^{-j\frac{\pi}{4}}, -0.9$	$0.9e^{-j\frac{\pi}{4}}, -0.7j$	$0.7e^{-j\frac{\pi}{4}}, 0.9$	$0.9e^{-j\frac{\pi}{4}}, 0.7j$	$0.8e^{-j\frac{\pi}{4}}, 0.8e^{-j\frac{pi}{4}}$

199 (an observed sequence is $N=200$ samples long). In other words, these coefficients satisfy the following relation:

$$1+\sum_{p=1}^{4} a_p^{(k)}[n]z^{-p} = \prod_{p=1}^{4}\left(1+\varrho_p^{(k)}[n]z^{-1}\right),$$

where

$$\varrho_p^{(k)}[n] = \varrho_p^{(k)}[0] + \frac{n}{N}\left(\varrho_p^{(k)}[N] - \varrho_p^{(k)}[0]\right) \quad (47)$$

with the initial and final values of the poles for each source specified in Table II.

In this experiment, we employ the "multiple experiments" framework, estimating the sources' covariance matrices from M independent realizations of the mixtures (each of length $N=200$). The covariance estimation is based on estimation of the TVAR parameters from the M independent realizations using the general Dym–Gohberg algorithm [1], [10] for TVAR parameters estimation. It is important to emphasize that we do *not* exploit the knowledge of our specific TVAR model (of linear drift of the poles) in the estimation process. The only information used is the TVAR order ($P=4$). The estimated TVAR parameters are obtained from the Dym–Gohberg algorithm as free parameters, with no particular temporal inter-relations between them.

In Fig. 4, we show empirical results for the total average ISR, given by

$$\mathrm{ISR} = \frac{1}{K(K-1)} \sum_{\substack{k,\ell=1\\ k\neq\ell}}^{K} \mathrm{ISR}_{k,\ell} \quad (48)$$

(with $K=5$ in this case), versus the number of experiments M. As in the previous experiment, we compare the results to the iCRLB (also averaged over all couples), to the ML performance (obtained using the true TVAR parameters) and to WASOBI.[5] Note that the iCRLB in this case is simply given by the single-experiment iCRLB divided by M (since the M experiments are independent).

Evidently, the ML performance nearly coincides with the iCRLB for all M. The BML performance approaches the iCRLB rather rapidly, and for $M>10$ the difference becomes smaller than 1 dB. Recall that for estimating arbitrary $N\times N$ covariance matrices, M generally has to be much larger than N; Nevertheless, with a TVAR model of order P it is sufficient to have $MN \gg PN$, namely $M \gg P$, even if M is much smaller than N—which is well-exploited in our case. As could

[5]Using a slightly modified version of the code in [20], adapted to estimating the time-lagged correlations from multiple snapshots.

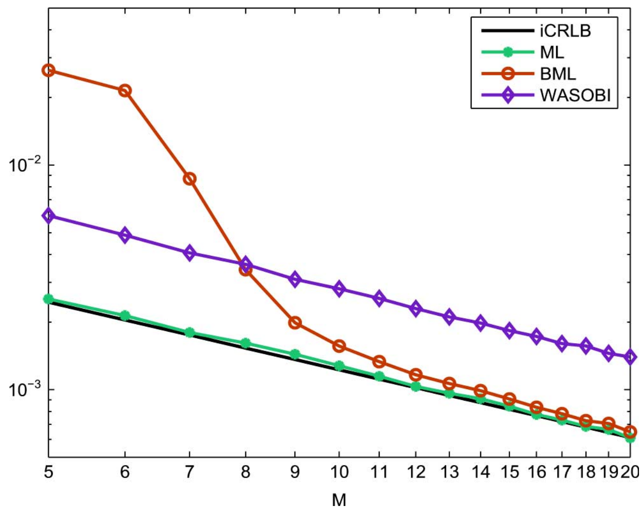

Fig. 4. Total average ISR versus M: iCRLB, ML, BML and WASOBI.

be expected, WASOBI, which attains the iCRLB for stationary AR sources, cannot exploit the temporal variability of the sources to the extent exploited by our ML and BML (and predicted by our iCRLB).

VI. Conclusion

We presented a general framework for optimal semi-blind and blind separation of Gaussian sources with arbitrary covariance structures. We derived the induced CRLB on the attainable ISR in terms of the sources' covariance matrices, developed the semi-blind ML separation and proposed an approach for fully blind (or "nearly" fully blind) ML separation. We demonstrated that this new framework enables, under various scenarios, to attain substantial performance gains with respect to conventional approaches (aimed at stationary sources) by proper exploitation of the covariance diversity of the sources.

Appendix
Invariance of the iCRLB

Let $\boldsymbol{S}$ denote a $K \times N$ matrix of source signals (not necessarily independent), whose joint distribution is known up to an unknown parameters vector $\boldsymbol{\theta}$, and let $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$ denote their observed mixtures. Let the vector of all unknown parameters be constructed as $\boldsymbol{\xi} \triangleq [\text{vec}^T(\boldsymbol{B})\ \boldsymbol{\theta}^T]^T$. Now consider a set of M statistically independent experiments, such that in each experiment the entire set $\boldsymbol{S}_m$ of K source signals is independently redrawn from the same joint distribution and mixed by the same mixing-matrix $\boldsymbol{A}$, such that the observed mixtures are given by $\boldsymbol{X}_m = \boldsymbol{A}\boldsymbol{S}_m$, $m = 1, \ldots, M$. Let $\boldsymbol{J}_{\boldsymbol{\xi}}(M)$ denote the FIM for estimation of $\boldsymbol{\xi}$ from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M$. Evidently (see, e.g., [11]), we have $\boldsymbol{J}_{\boldsymbol{\xi}}(M) = M \cdot \boldsymbol{J}_{\boldsymbol{\xi}}(1)$.

Now let $\hat{\boldsymbol{\xi}}_{\text{ML}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ denote the ML estimate of $\boldsymbol{\xi}$ based on the M independent experiments. Asymptotically (as $M \rightarrow \infty$) the ML estimate attains the CRLB, namely, $\hat{\boldsymbol{\xi}}_{\text{ML}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ is asymptotically unbiased, and its covariance coincides asymptotically with $\boldsymbol{J}_{\boldsymbol{\xi}}^{-1}(M)$.

Let us show now that $\hat{\boldsymbol{B}}_{\text{ML}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ is *ISR-equivariant*. Our proof essentially resembles the one presented in [6]; however, in [6] all sources are assumed to have an i.i.d. time-structure, and their marginal distributions are implicitly assumed to be known. In the following, we address our slightly more general model.

Theorem 1: The ML estimate of $\boldsymbol{B}$ from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M$, denoted $\hat{\boldsymbol{B}}_{\text{ML}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ is an *ISR-equivariant* estimator, namely, for any $K \times K$ nonsingular matrix $\boldsymbol{Q}$, $\hat{\boldsymbol{B}}_{\text{ML}}(\boldsymbol{Q}\boldsymbol{X}_1, \ldots, \boldsymbol{Q}\boldsymbol{X}_M) = \hat{\boldsymbol{B}}_{\text{ML}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M) \cdot \boldsymbol{Q}^{-1}$

Proof: We use the vectorized version of the mixing model $\boldsymbol{X}_m = \boldsymbol{A}\boldsymbol{S}_m = \boldsymbol{B}^{-1}\boldsymbol{S}_m$ as

$$\boldsymbol{x}_m = (\boldsymbol{B}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{s}_m, \quad m = 1, \ldots, M \tag{49}$$

where $\boldsymbol{s}_m = \text{vec}(\boldsymbol{S}_m^T)$ and $\boldsymbol{x}_m = \text{vec}(\boldsymbol{X}_m^T)$. This implies

$$f_x(\boldsymbol{x}; \boldsymbol{B}, \boldsymbol{\theta}) = \frac{1}{\det(\boldsymbol{B}^{-1} \otimes \boldsymbol{I}_N)} f_s((\boldsymbol{B} \otimes \boldsymbol{I}_N)\boldsymbol{x}; \boldsymbol{\theta}) \tag{50}$$

where $f_x(\cdot; \boldsymbol{B}, \boldsymbol{\theta})$ and $f_s(\cdot; \boldsymbol{\theta})$ denote the *pdf*-s of each of the random vectors $\boldsymbol{x}_m$ and $\boldsymbol{s}_m$ (respectively). With slight modification of notations, using the relation $\det(\boldsymbol{B}^{-1} \otimes \boldsymbol{I}_N) = (\det \boldsymbol{B})^{-N}$ and exploiting the statistical independence of the M experiments, we get

$$\begin{aligned} &f_{x(1:M)}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M; \boldsymbol{B}, \boldsymbol{\theta}) \\ &\quad = \prod_{m=1}^{M} f_x(\boldsymbol{X}_m; \boldsymbol{B}, \boldsymbol{\theta}) = \prod_{m=1}^{M} (\det \boldsymbol{B})^N f_s(\boldsymbol{B}\boldsymbol{X}_m; \boldsymbol{\theta}) \end{aligned} \tag{51}$$

where $f_{x(1:M)}(\cdot; \boldsymbol{B}, \boldsymbol{\theta})$ denotes the joint *pdf* of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M$.

Let $\hat{\boldsymbol{B}}_0$ and $\hat{\boldsymbol{\theta}}_0$ denote the ML estimates of $\boldsymbol{B}$ and $\boldsymbol{\theta}$ obtained with given measurements $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M$. This implies the inequality

$$\begin{aligned} &f_{x(1:M)}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M; \hat{\boldsymbol{B}}_0, \hat{\boldsymbol{\theta}}_0) \\ &\qquad \geq f_{x(1:M)}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M; \tilde{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}) \end{aligned} \tag{52}$$

for all nonsingular $\tilde{\boldsymbol{B}}$ and $\hat{\boldsymbol{\theta}}$, which in turn implies (using (51))

$$\begin{aligned} &(\det \hat{\boldsymbol{B}}_0)^{NM} \prod_{m=1}^{M} f_s(\hat{\boldsymbol{B}}_0\boldsymbol{X}_m; \hat{\boldsymbol{\theta}}_0) \\ &\qquad \geq (\det \tilde{\boldsymbol{B}})^{NM} \prod_{m=1}^{M} f_s(\tilde{\boldsymbol{B}}\boldsymbol{X}_m; \hat{\boldsymbol{\theta}}). \end{aligned} \tag{53}$$

Now let $\boldsymbol{Q}$ denote an arbitrary $K \times K$ nonsingular matrix. We observe (using (51) and (53)) that for all nonsingular $\tilde{\boldsymbol{B}}$ and $\hat{\boldsymbol{\theta}}$

$$\begin{aligned} &f_{x(1:M)}(\boldsymbol{Q}\boldsymbol{X}_1, \ldots, \boldsymbol{Q}\boldsymbol{X}_M; \hat{\boldsymbol{B}}_0\boldsymbol{Q}^{-1}, \hat{\boldsymbol{\theta}}_0) \\ &\quad = (\det \hat{\boldsymbol{B}}_0)^{NM} (\det \boldsymbol{Q})^{-NM} \prod_{m=1}^{M} f_s(\hat{\boldsymbol{B}}_0\boldsymbol{X}_m; \hat{\boldsymbol{\theta}}_0) \\ &\quad \geq (\det \tilde{\boldsymbol{B}})^{NM} (\det \boldsymbol{Q})^{-NM} \prod_{m=1}^{M} f_s(\tilde{\boldsymbol{B}}\boldsymbol{X}_m; \hat{\boldsymbol{\theta}}). \end{aligned} \tag{54}$$

Now let $\tilde{\boldsymbol{B}} = \hat{\boldsymbol{B}}\boldsymbol{Q}$ where $\hat{\boldsymbol{B}} = \tilde{\boldsymbol{B}}\boldsymbol{Q}^{-1}$ is any nonsingular matrix. We have

$$\begin{aligned} &f_{x(1:M)}(\boldsymbol{Q}\boldsymbol{X}_1, \ldots \boldsymbol{Q}\boldsymbol{X}_M; \hat{\boldsymbol{B}}, \hat{\boldsymbol{\theta}}) \\ &\quad = f_{x(1:M)}(\boldsymbol{Q}\boldsymbol{X}_1, \ldots, \boldsymbol{Q}\boldsymbol{X}_M; \tilde{\boldsymbol{B}}\boldsymbol{Q}^{-1}, \hat{\boldsymbol{\theta}}) \\ &\quad = (\det \tilde{\boldsymbol{B}})^{NM} (\det \boldsymbol{Q})^{-NM} \prod_{m=1}^{M} f_s(\tilde{\boldsymbol{B}}\boldsymbol{X}_m; \hat{\boldsymbol{\theta}}), \end{aligned} \tag{55}$$

which coincides with the right-hand side of (54). We therefore conclude (using (53) and (55)) that

$$f_{x(1:M)}(\boldsymbol{QX}_1,\ldots,\boldsymbol{QX}_M;\hat{\boldsymbol{B}}_0\boldsymbol{Q}^{-1},\hat{\boldsymbol{\theta}}_0) \geq f_{x(1:M)}(\boldsymbol{QX}_1,\ldots,\boldsymbol{QX}_M;\hat{\boldsymbol{B}},\hat{\boldsymbol{\theta}}) \quad (56)$$

for all nonsingular $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{\theta}}$. In other words, when the observations $\boldsymbol{X}_m$ are changed into $\boldsymbol{QX}_m$, the ML estimate $\hat{\boldsymbol{B}}_0$ changes into $\hat{\boldsymbol{B}}_0\boldsymbol{Q}^{-1}$ (and the ML estimate $\hat{\boldsymbol{\theta}}_0$ of $\boldsymbol{\theta}$ remains unchanged). ■

Thus, since the MSE of $\hat{\boldsymbol{B}}_{\mathrm{ML}}(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_M)$ coincides (asymptotically) with the CRLB, this implies that the asymptotic iCRLB is also invariant w.r.t. $\boldsymbol{A}$. Furthermore, since, as discussed in Section II, the iCRLB is a linear transformation of the CRLB (on $\boldsymbol{B}$), and since the single-experiment CRLB is given by $\boldsymbol{J}_{\boldsymbol{\xi}}^{-1}(1) = M \cdot \boldsymbol{J}_{\boldsymbol{\xi}}^{-1}(M)$, it follows that the single-experiment iCRLB also differs from the M-experiments iCRLB only by a factor of M, and must therefore be invariant w.r.t. $\boldsymbol{A}$ as well.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers, and particularly Reviewer 2, for suggesting ideas which were helpful in simplifying some of the derivations.

REFERENCES

[1] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1298–1311, Apr. 2007.

[2] B. Afsari, "Simple LU and QR based non-orthogonal matrix joint diagonalization," in *Lecture Notes in Computer Science (LNCS 3889): Proc. ICA*, 2006, pp. 17–24.

[3] S.-I. Amari and J.-F. Cardoso, "Blind source separation-semiparametric statistical approach," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2692–2700, Nov. 1997.

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

[5] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO*, 1994, pp. 776–779.

[6] J.-F. Cardoso, "The invariant approach to source separation," in *Proc., NOLTA'95*, 1995.

[7] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017–3030, 1996.

[8] S. Dégerine and A. Zaïdi, "Separation of an instantaneous mixture of Gaussian autoregressive sources by the exact maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1499–1512, Jun. 2004.

[9] E. Doron, A. Yeredor, and P. Tichavský, "Cramér-Rao lower bound for blind separation of stationary parametric Gaussian sources," *IEEE Signal Process. Lett.*, vol. 14, no. 6, pp. 417–420, Jun. 2007.

[10] H. Dym and I. Gohberg, "Extensions of band matrices with band inverses," *Linear Algebra Appl.*, vol. 36, pp. 1–24, 1981.

[11] S. M. Kay*, Fundamentals of Statistical SIgnal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[12] E. Ollila, K. Hyon-Jung, and V. Koivunen, "Compact Cramér–Rao bound expression for independent component analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1421–1428, Apr. 2008.

[13] E. Ollila, H. Oja, and V. Koivunen, "Complex-valued ICA based on a pair of generalized covariance matrices," *Comput. Stat. Data Anal.*, vol. 52, no. 7, pp. 3789–3805, 2008.

[14] D.-T. Pham, "Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion," *Signal Process.*, vol. 81, pp. 855–870, 2001.

[15] D.-T. Pham, "Joint approximate diagonalization of positive definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, pp. 1136–1152, 2001.

[16] D.-T. Pham, "Blind separation of cyclostationary sources using joint block approximate diagonalization," in *Lecture Notes in Computer Science (LNCS 4666): Proc. ICA*, 2007, pp. 244–251.

[17] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of non-stationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.

[18] D.-T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, 1997.

[19] A. Souloumiac, "Nonorthogonal joint diagonalization by combining givens and hyperbolic rotations," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2222–2231, Jun. 2009.

[20] P. Tichavský, Matlab Code for WEDGE and WASOBI. 2009 [Online]. Available: http://si.utia.cas.cz/Tichavsky.html

[21] P. Tichavský, E. Doron, A. Yeredor, and J. Nielsen, "A computationally affordable implementation of an asymptotically optimal BSS algorithm for AR sources," in *Proc. EUSIPCO*, 2006.

[22] P. Tichavský, Z. Koldovsky, and E. Oja, "Performance analysis of the fastICA algorithm and Cramér-Rao bounds for linear independent component analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.

[23] P. Tichavský and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 878–891, Apr. 2009.

[24] P. Tichavský, A. Yeredor, and Z. Koldobský, "A fast asymptotically efficient algorithm for blind separation of a linear mixture of block-wise stationary autoregressive processes," in *Proc. ICASSP*, 2009.

[25] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3270–3278, Sep. 2006.

[26] A. Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *IEEE Signal Process. Lett.*, vol. 7, no. 7, pp. 197–200, Jul. 2000.

[27] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.

[28] A. Yeredor, "On hybrid exact-approximate joint diagonalization," in *Proc. 3rd Int. Workshop Computational Advances Multi-Sensor Adaptive Processing (CAMSAP)*, 2009.

Arie Yeredor (M'99–SM'02) received the B.Sc. (*summa cum laude*) and Ph.D. degrees in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1984 and 1997, respectively.

He is currently with the School of Electrical Engineering, Department of Electrical Engineering-Systems, TAU, where his research and teaching areas are in statistical and digital signal processing. He also holds a consulting position with NICE Systems, Inc., Ra'anana, Israel, in the fields of speech and audio processing, video processing, and emitter location algorithms.

Dr. Yeredor previously served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II, and he is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He served as Technical Co-Chair of The Third International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) 2009. He has been awarded the yearly Best Lecturer of the Faculty of Engineering Award (at TAU) six times. He is a member of the IEEE Signal Processing Society's Signal Processing Theory and Methods (SPTM) Technical Committee.