A MAXIMUM LIKELIHOOD "IDENTIFICATION-CORRECTION" SCHEME OF SUB-OPTIMAL "SEDJOCO" SOLUTIONS FOR SEMI-BLIND SOURCE SEPARATION

Amir Weiss[†], *Arie Yeredor*[†], *Sher Ali Cheema*[⋆] *and Martin Haardt*[⋆]

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

[⋆]Communications Research Laboratory
Ilmenau University of Technology
P.O. Box 100565, D-98684 Ilmenau, Germany
{sher-ali.cheema, martin.haardt}@tu-ilmenau.de

ABSTRACT

The "Sequentially Drilled" Joint Congruence (SeDJoCo) transformation is a set of matrix transformation equations, which coincide with the Likelihood Equations for semi-blind source separation, when each source is modeled as a zero-mean Gaussian process with a known (and distinct) temporal covariance matrix. Therefore, with such a model a solution of SeDJoCo can lead to the Maximum Likelihood (ML) estimate of the separating matrix, which is asymptotically optimal. However, as we have shown in previous work, multiple solutions of SeDJoCo may exist, and the selection of the optimal solution among these (corresponding to the global maximum of the likelihood function) is therefore of considerable interest. In this paper we further extend our results by proposing a new ML approach for the identification and correction of a sub-optimal solution, assuming sources of unrestricted, general temporal covariance structures. We demonstrate the resulting improvement in simulation with non-stationary sources.

Index Terms— Maximum Likelihood, Semi-Blind Source Separation, SeDJoCo, Permutation

1. INTRODUCTION

Traditionally, Blind Source Separation (BSS) and Independent Component Analysis (ICA) employ a general model, free of *a-priori* statistical information regarding the sources, except for their mutual statistical independence. Consequently, a Maximum Likelihood (ML) approach for the separation cannot be applied with such models. Quasi ML (QML) approaches (e.g., [1–3]), which make general assumptions (e.g., stationarity) on the sources and use an "educated guess" for the associated parameters (or spectra) have been proposed in different contexts. However, in some cases, commonly referred to as "semi-blind", statistical information on the sources is available, enabling to take a true ML approach and thereby to benefit from its asymptotic optimality [4] in the sense of minimal interference-to-source ratio (ISR).

The authors gratefully acknowledge the financial support by the German-Israeli Foundation (GIF), grant number I-1282-406.10/2014. The first author would like to thank the Yitzhak and Chaya Weinstein Research Institute for Signal Processing for a fellowship.

One of the prevalent models for semi-blind source separation (semi-BSS) is that of zero-mean Gaussian (possibly non-stationary) sources with known (and distinct) temporal covariance matrices. It has been shown (e.g., [1, 4] and [5] (chapter 7)) that for obtaining the ML estimate of the separating matrix in such models, i.e., solve the "Likelihood Equations", a special form of joint matrix transformations (reminiscent of, but essentially different from, approximate joint diagonalization) needs to be solved. That form was termed a "Sequentially Drilled" Joint Congruence (SeDJoCo) transformation in [6] (where it was also shown to be relevant in the context of Coordinated Beamforming (CBF), see also [7]).

Several solution strategies for SeDJoCo, all in the form of iterative algorithms, have been proposed in recent years, either explicitly or implicitly - see, e.g., QML [1], Iterative Relaxations [8] or Newton-Conjugate-Gradient (NCG) [6]. However, although the number of unknowns equals the number of equations in SeDJoCo, these equations are nonlinear, and therefore the solution is generally non-unique. Thus, these algorithms are all capable of finding an exact solution of the Likelihood Equations, which is guaranteed to be a stationary point of the Likelihood, but not necessarily its global maximizer, i.e., the true ML estimate.

Indeed, it was recently shown in [9] that when a solution of SeDJoCo exists (which is guaranteed under some mild conditions, see [6]), there exist at least $K!$ solutions (where K is the number of sources). These solutions correspond to different estimates of the separating matrix, involving all $K!$ possible permutations (and their respective induced scales) of the sources, where only one of these is the ML estimate (and the others are essentially different, i.e., differ by more than their permutation and scaling, as we shall explain in the sequel). We refer to the other $K! - 1$ (non-ML) solutions as *sub-optimal solutions*, since they usually still provide reasonable (but not optimal) separation in terms of ISR. It is important to realize, that in semi-BSS, where the sources' statistical properties are known (and distinct, such as in the abovementioned Gaussian model), there is no reason for scale or permutation ambiguity in the separation, because each reconstructed source can be matched to its own (known) model.

Thus, an intuitive heuristic approach for identifying such sub-optimal solutions and using them in computation of the ML solution was proposed in [9] for the case of stationary sources. In this work we propose a more rigorous and general (and approximately optimal) approach, which is applicable not only to stationary sources, but also to any Gaussian sources with general temporal covariance matrices.

2. PRELIMINARIES: SEDJOCO AND ITS MULTIPLE SOLUTIONS

For completeness of the presentation, we briefly review the problem formulation, the SeDJoCo set of equations and the characterization of its multiple solutions.

Consider the classic linear mixture model $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$, where $\boldsymbol{A} \in \mathbb{R}^{K\times K}$ is an unknown, deterministic (invertible) mixing-matrix, $\boldsymbol{S} = [\boldsymbol{s}_1\ \boldsymbol{s}_2\ \cdots\ \boldsymbol{s}_K]^{\mathrm{T}} \in \mathbb{R}^{K\times T}$ is an unknown sources' matrix of K statistically independent source signals, each of length T, and $\boldsymbol{X} \in \mathbb{R}^{K\times T}$ is the observed mixtures matrix, from which it is desired to estimate the demixing-matrix $\boldsymbol{B} \triangleq \boldsymbol{A}^{-1}$ and, subsequently, to which the estimated $\widehat{\boldsymbol{B}}$ will be applied in order to separate (estimate) the source signals. As shown in [1, 4] and [5] (chapter 7), when the source signals are zero-mean Gaussian, each with a known temporal covariance matrix $\boldsymbol{C}_k \triangleq E\left[\boldsymbol{s}_k\boldsymbol{s}_k^{\mathrm{T}}\right] \in \mathbb{R}^{T\times T}$ distinct from all other covariance matrices, the ML estimate $\widehat{\boldsymbol{B}}_{\mathrm{ML}}$, corresponding to the global maximum of the likelihood, is a solution of the following set of K^2 equations (K vector equations with K elements each):

$$\widehat{\boldsymbol{B}}\boldsymbol{Q}_k\widehat{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{e}_k = \boldsymbol{e}_k, \quad \forall k \in \{1,\ldots,K\}, \tag{1}$$

where the "pinning vector" $\boldsymbol{e}_k$ denotes the k'th column of the $K\times K$ identity matrix, and where the ordered set of matrices

$$\boldsymbol{Q}_k \triangleq \frac{1}{T}\boldsymbol{X}\boldsymbol{C}_k^{-1}\boldsymbol{X}^{\mathrm{T}} \in \mathbb{R}^{K\times K}, \quad \forall k \in \{1,\ldots,K\} \tag{2}$$

are termed the "target-matrices". The solution $\widehat{\boldsymbol{B}}$ of (1) jointly transforms the set of "target-matrices" so that the k'th column (and, by symmetry of $\boldsymbol{Q}_k$, row) of the transformed matrix $\widehat{\boldsymbol{B}}\boldsymbol{Q}_k\widehat{\boldsymbol{B}}^{\mathrm{T}}$ equals the vector $\boldsymbol{e}_k$. This transformation is illustrated in Fig. 1 for $K = 3$.

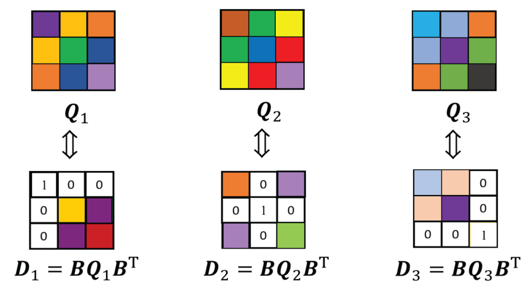

Fig. 1: Illustration of SeDJoCo as a joint matrix transformation for $K = 3$.

Under the conditions stated above, a solution $\widehat{\boldsymbol{B}}$ always exists, but is not unique, as at least $K! - 1$ other essentially different (for $K > 2$) solutions exist. To show that, consider a SeDJoCo problem with a given set $\{\boldsymbol{Q}_k\}_{k=1}^{K}$. Additionally, consider yet another (different) SeDJoCo problem, with the same set of target matrices, but in a slightly different order, in which the first two are swapped: $\{\widetilde{\boldsymbol{Q}}_k\}_{k=1}^{K}$, defined as $\widetilde{\boldsymbol{Q}}_1 = \boldsymbol{Q}_2, \widetilde{\boldsymbol{Q}}_2 = \boldsymbol{Q}_1$, and $\widetilde{\boldsymbol{Q}}_k = \boldsymbol{Q}_k,\ \forall k \in \{3,\ldots,K\}$.

Let $\boldsymbol{B}$ denote a solution for the latter. Now define $\boldsymbol{P}_{1,2}$ as the (symmetric) permutation matrix that swaps the first and second elements of a vector, namely $\boldsymbol{P}_{1,2}\boldsymbol{e}_1 = \boldsymbol{e}_2$, $\boldsymbol{P}_{1,2}\boldsymbol{e}_2 = \boldsymbol{e}_1$ and $\boldsymbol{P}_{1,2}\boldsymbol{e}_k = \boldsymbol{e}_k$ for all other $k \in \{3,\ldots,K\}$. We claim that the matrix $\boldsymbol{B}' = \boldsymbol{P}_{1,2}\boldsymbol{B}$ is a solution for the original SeDJoCo problem with $\{\boldsymbol{Q}_k\}_{k=1}^{K}$, since

$$\begin{aligned} \boldsymbol{B}'\boldsymbol{Q}_1\boldsymbol{B}'^{\mathrm{T}}\boldsymbol{e}_1 &= \boldsymbol{P}_{1,2}\boldsymbol{B}\boldsymbol{Q}_1\boldsymbol{B}^{\mathrm{T}}\boldsymbol{P}_{1,2}^{\mathrm{T}}\boldsymbol{e}_1 \\ &= \boldsymbol{P}_{1,2}\boldsymbol{B}\widetilde{\boldsymbol{Q}}_2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{e}_2 = \boldsymbol{P}_{1,2}\boldsymbol{e}_2 = \boldsymbol{e}_1; \end{aligned} \tag{3}$$

similarly, $\boldsymbol{B}'\boldsymbol{Q}_2\boldsymbol{B}'^{\mathrm{T}}\boldsymbol{e}_2 = \boldsymbol{e}_2$; and, trivially, $\boldsymbol{B}'\boldsymbol{Q}_k\boldsymbol{B}'^{\mathrm{T}}\boldsymbol{e}_k = \boldsymbol{e}_k$ for all other $k \in \{3,\ldots,K\}$. Note that generally, $\boldsymbol{B}'$ would be essentially different from other solutions of the original problem, namely the difference will not merely be up to some permutations of rows and / or columns, because for the SeDJoCo problem with the permuted target matrices, the entire set of implied polynomial equations (in the elements of $\boldsymbol{B}$) is changed (with respect to the set implied by the original problem), in the sense that different coefficients multiply different second-order monomials in the set.

Since any permutation matrix can be expressed as the product of two-elements-permutation matrices, the above result may be generalized and we conclude that the number of solutions having this particular structure is equal to the number of possible permutations, $K!$. Of course, one of these solution will (most probably) be the desired ML solution. The set of $K! - 1$ other solutions in this "family", denoted by $\mathcal{P}$, are referred to as the sub-optimal, "wrongly permuted" solutions. We now proceed to describe the proposed method for identification and correction of such sub-optimal solutions.

3. THE PROPOSED "IDENTIFICATION-CORRECTION" SCHEME

Let $\widehat{\boldsymbol{B}}$ be a solution of (1) for the Gaussian semi-BSS problem described in the previous section. We remind that since the covariance matrices $\{\boldsymbol{C}_k\}_{k=1}^{K}$ are known, the ML solution should be permutation-free and properly scaled in accordance with this prior available information (in contrast to a general BSS problem). Therefore, when $\widehat{\boldsymbol{B}}$ is the ML estimate, the estimated sources, given by the rows of $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{B}}\boldsymbol{X}$, should be ordered correctly, and, under mild asymptotic conditions, have approximately the same covariance matrices as the true sources. However, if $\widehat{\boldsymbol{B}} \in \mathcal{P}$, the estimated sources may be wrongly permuted and scaled. In the following, we shall assume that if a non-ML solution has been obtained, it is a sub-optimal solution belonging to $\mathcal{P}$.

When $\widehat{\boldsymbol{B}}$ is fixed, each row of $\widehat{\boldsymbol{S}}$ is obviously (zero-mean) Gaussian. Denote by $\widehat{\boldsymbol{C}}_k \triangleq E\left[\hat{\boldsymbol{s}}_k \hat{\boldsymbol{s}}_k^{\mathrm{T}}\right] \in \mathbb{R}^{T\times T}$ (for all $k \in \{1,\ldots,K\}$) as the covariance matrices of the estimated sources (assuming that $\widehat{\boldsymbol{B}}$ is fixed). If $\widehat{\boldsymbol{B}} \in \mathcal{P}$, we have

$$\hat{\boldsymbol{s}}_k \approx \gamma_k \boldsymbol{s}_{p(k)}, \quad \forall k \in \{1,\ldots,K\}, \tag{4}$$

where $\{\gamma_k \in \mathbb{R}\}_{k=1}^K$ are unknown (deterministic) scaling factors and $p(\cdot) : \{1,\ldots,K\} \rightarrow \{1,\ldots,K\}$ is an unknown (deterministic and injective) permutation function. Thus,

$$\widehat{\boldsymbol{C}}_k \approx \gamma_k^2 \cdot \boldsymbol{C}_{p(k)}, \quad \forall k \in \{1,\ldots,K\}. \tag{5}$$

Assuming that the estimated sources are approximately statistically independent (given $\widehat{\boldsymbol{B}}$), their joint pdf is given by (to simplify the notation, we use $\boldsymbol{\gamma} \triangleq [\gamma_1\ \gamma_2\ \ldots\ \gamma_K]^{\mathrm{T}}$):

$$f_{\widehat{\boldsymbol{S}}}(\widehat{\boldsymbol{S}};\boldsymbol{\gamma},p(\cdot)) \approx \prod_{k=1}^{K} \left|2\pi\widehat{\boldsymbol{C}}_k\right|^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\hat{\boldsymbol{s}}_k^{\mathrm{T}}\widehat{\boldsymbol{C}}_k^{-1}\hat{\boldsymbol{s}}_k\right), \tag{6}$$

where $|\cdot|$ denotes the determinant. Taking the log, substituting (5) and dropping irrelevant constants yield the log-likelihood

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\gamma},p(\cdot)\right) &\triangleq -\frac{1}{2}\sum_{k=1}^{K}\left\{\log\left(\gamma_k^{2T}\left|\boldsymbol{C}_{p(k)}\right|\right) + \frac{1}{\gamma_k^2}\hat{\boldsymbol{s}}_k^{\mathrm{T}}\boldsymbol{C}_{p(k)}^{-1}\hat{\boldsymbol{s}}_k\right\}\\ &= c - \frac{1}{2}\sum_{k=1}^{K}\left\{T\log\gamma_k^2 + \frac{1}{\gamma_k^2}\hat{\boldsymbol{s}}_k^{\mathrm{T}}\boldsymbol{C}_{p(k)}^{-1}\hat{\boldsymbol{s}}_k\right\}, \end{aligned}\tag{7}$$

where c is a constant independent of $p(\cdot)$. Accordingly, the ML estimates of $\boldsymbol{\gamma}$ and $p(\cdot)$ are given by

$$\{\hat{\boldsymbol{\gamma}}_{\text{ML}}, \hat{p}_{\text{ML}}(\cdot)\} = \underset{\boldsymbol{\gamma}\in\mathbb{R}^K, p(\cdot)\in\Pi}{\operatorname{argmin}} F\left(\boldsymbol{\gamma},p(\cdot)\right), \tag{8}$$

where Π is the set of all valid permutation functions (with cardinality $|\Pi| = K!$), and where

$$F\left(\boldsymbol{\gamma},p(\cdot)\right) \triangleq \frac{1}{2}\sum_{k=1}^{K}\left\{T\log\gamma_k^2 + \frac{1}{\gamma_k^2}\hat{\boldsymbol{s}}_k^{\mathrm{T}}\boldsymbol{C}_{p(k)}^{-1}\hat{\boldsymbol{s}}_k\right\}. \tag{9}$$

Clearly, for any fixed $p(\cdot)$, the (candidate) optimal γ_ℓ is given by differentiating $F\left(\boldsymbol{\gamma},p(\cdot)\right)$ and equating the derivative to zero. Indeed, by doing so we have

$$\frac{\partial}{\partial\gamma_\ell}F\left(\boldsymbol{\gamma},p(\cdot)\right) = \frac{T}{\gamma_\ell} - \frac{1}{\gamma_\ell^3}\hat{\boldsymbol{s}}_\ell^{\mathrm{T}}\boldsymbol{C}_{p(\ell)}^{-1}\hat{\boldsymbol{s}}_\ell \stackrel{!}{=} 0, \tag{10}$$

yielding the (candidate) ML estimates of the scaling factors

$$\hat{\gamma}_{\ell_{\text{ML}}}^2 = \frac{1}{T}\hat{\boldsymbol{s}}_\ell^{\mathrm{T}}\boldsymbol{C}_{p(\ell)}^{-1}\hat{\boldsymbol{s}}_\ell, \quad \forall\ell \in \{1,\ldots,K\}, \tag{11}$$

up to an inevitable sign ambiguity. Notice that we then get

$$F\left(\hat{\boldsymbol{\gamma}}_{\text{ML}},p(\cdot)\right) = \frac{T}{2}\sum_{k=1}^{K}\left\{\log\hat{\gamma}_{k_{\text{ML}}}^2 + 1\right\}, \tag{12}$$

so that in order to determine the minimizer of $F\left(\boldsymbol{\gamma},p(\cdot)\right)$, it is sufficient to only minimize $\sum_{k=1}^{K}\log\hat{\gamma}_{k_{\text{ML}}}^2$. This enables us to efficiently find the ML estimates of $\boldsymbol{\gamma}$ and $p(\cdot)$ as follows. Given the estimated sources $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{B}}_0\boldsymbol{X}$, compute a $K\times K$ "confusion matrix" $\boldsymbol{G}$, whose (k,ℓ)'th element is defined as

$$G_{(k,\ell)} = \log\left(\hat{\boldsymbol{s}}_\ell^{\mathrm{T}}\boldsymbol{C}_k^{-1}\hat{\boldsymbol{s}}_\ell/T\right). \tag{13}$$

The ML estimate of $p(\cdot)$ is then given by the permutation which re-orders the rows of $\boldsymbol{G}$ so as to minimize its trace,

$$\widehat{\boldsymbol{P}}_{\text{ML}} = \underset{\widehat{\boldsymbol{P}}\in\boldsymbol{\Pi}}{\operatorname{argmin}}\ \text{Trace}\left(\widehat{\boldsymbol{P}}\boldsymbol{G}\right), \tag{14}$$

where $\widehat{\boldsymbol{P}}_{\text{ML}}$ and $\boldsymbol{\Pi}$ are matrix representations of $\hat{p}_{\text{ML}}(\cdot)$ and of the set Π, respectively. Subsequently

$$\hat{\boldsymbol{\gamma}}_{\text{ML}} = \exp\left(0.5\cdot\text{diag}\left(\widehat{\boldsymbol{P}}_{\text{ML}}\boldsymbol{G}\right)\right), \tag{15}$$

where $\text{diag}(\cdot)$ forms a vector of the diagonal elements of its argument, and $\exp(\cdot)$ works elementwise.

Given this ML estimate of the "wrong" permutation and scaling in $\widehat{\boldsymbol{B}}$, it is readily possible to modify $\widehat{\boldsymbol{B}}$ so as to "undo" these permutation and scaling. Of course, the resulting modified $\widehat{\boldsymbol{B}}$ will no longer be a solution of the given SeDJoCo problem (of which $\widehat{\boldsymbol{B}}$ was). However, it is reasonable to assume that re-initializing a gradient-based iterative solution algorithm with the modified $\widehat{\boldsymbol{B}}$ would promote convergence to a nearby solution (in the same family) that would be permutation- and scaling-free - namely the ML solution.

Thus, the ML "identification-correction" scheme for a given solution $\widehat{\boldsymbol{B}}_0$ of (1) takes the following steps:

1. Compute the estimated source matrix $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{B}}_0\boldsymbol{X}$;
2. Compute $\boldsymbol{G}$ using (13) and then find $\widehat{\boldsymbol{P}}_{\text{ML}}$ and $\hat{\boldsymbol{\gamma}}_{\text{ML}}$;
3. Compute a new initial guess $\boldsymbol{B}_{\text{init}} = \widehat{\boldsymbol{P}}_{\text{ML}}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}}_{\text{ML}}^{-1}\widehat{\boldsymbol{B}}_0$;
4. Solve (1) (e.g., with NCG [9]) using the new initial guess $\boldsymbol{B}_{\text{init}}$.

$\widehat{\boldsymbol{\Gamma}}_{\text{ML}}$ above is a diagonal matrix with $\hat{\gamma}_{k_{\text{ML}}}$ as its (k,k)'th element; $\widehat{\boldsymbol{P}}_{\text{ML}}$ (in (14)) can either be found precisely by an exhaustive search (for small values of K), or approximated by a greedy search (for larger values of K). Note that when $\widehat{\boldsymbol{B}}_0$ happens to be the ML solution, we would get $\widehat{\boldsymbol{P}}_{\text{ML}} = \boldsymbol{I}$ and $\widehat{\boldsymbol{\Gamma}}_{\text{ML}} \approx \boldsymbol{I}$, so $\boldsymbol{B}_{\text{init}}$ would be (almost) identical to $\widehat{\boldsymbol{B}}_0$, and the iterative solution would converge back to $\widehat{\boldsymbol{B}}_0$. The scheme can thus be used to identify an optimal solution: If $\widehat{\boldsymbol{P}}_{\text{ML}} \neq \boldsymbol{I}$ then $\widehat{\boldsymbol{B}}_0$ is sub-optimal, but if $\widehat{\boldsymbol{P}}_{\text{ML}} = \boldsymbol{I}$ then $\widehat{\boldsymbol{B}}_0$ is very likely to be optimal, and no further action is required.

It is interesting to compare our approach with the one presented in [9] for the case of stationary sources: Exploiting the ensuing Toeplitz structure of the covariance matrices $\boldsymbol{C}_k$, and the fact that Toeplitz matrices are asymptotically diagonalized by the (normalized) DFT matrix $\boldsymbol{F} \in \mathbb{C}^{T\times T}$, we obtain

$$\begin{aligned}\hat{\gamma}^2_{k_{\mathrm{ML}}} &= \frac{1}{T}\hat{\boldsymbol{s}}_k^{\mathrm{T}}\boldsymbol{C}_{p(k)}^{-1}\hat{\boldsymbol{s}}_k = \frac{1}{T}\hat{\boldsymbol{s}}_k^{\mathrm{T}}\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{C}_{p(k)}^{-1}\boldsymbol{F}^H\boldsymbol{F}\hat{\boldsymbol{s}}_k \\ &= \frac{1}{T}\left(\boldsymbol{F}\hat{\boldsymbol{s}}_k\right)^H\left(\boldsymbol{F}\boldsymbol{C}_{p(k)}\boldsymbol{F}^H\right)^{-1}\left(\boldsymbol{F}\hat{\boldsymbol{s}}_k\right) \\ &\triangleq \frac{1}{T}\tilde{\hat{\boldsymbol{s}}}_k^H\widetilde{\boldsymbol{C}}_{p(k)}^{-1}\tilde{\hat{\boldsymbol{s}}}_k \approx \frac{1}{2\pi}\int_0^{2\pi}\frac{\hat{P}_k(\omega)}{P_{p(k)}(\omega)}\mathrm{d}\omega. \qquad (16)\end{aligned}$$

Here $\tilde{\hat{\boldsymbol{s}}}_k$ is the DFT of $\hat{\boldsymbol{s}}_k$, $\widetilde{\boldsymbol{C}}_{p(k)}$ is diagonal, $\hat{P}_k(\omega)$ is the periodogram spectrum estimate of $\hat{\boldsymbol{s}}_k$ and $P_k(\omega)$ is the k'th source's true spectrum. Thus, the "closeness" measure of each estimated source to its true spectrum, which was a heuristic least squares measure (between the scaled estimated spectrum and the true spectrum) in [9], is replaced here with a measure that integrates over the ratio between the periodogram and the true spectrum.

Next, we demonstrate the resulting near-optimal performance, in the sense of attaining the induced Cramér-Rao lower bound (iCRLB, [4]) on the ISR.

4. SIMULATION RESULTS

We mixed $K = 8$ Gaussian non-stationary sources, obtained by first generating stationary Gaussian Moving Average (MA) processes of order 6 (using 8 different fixed random length-6 FIR filters), and then multiplying each process by an "envelope signal" given by $e_k[n] \triangleq \sigma_k + \cos\left(2\pi n/N_k + \phi_k\right)$, where $\{\phi_k, N_k, \sigma_k\}_{k=1}^{8}$ are fixed random values. In each trial the mixing matrix' elements were drawn independently from a standard Gaussian distribution, and we used T samples of the mixture signals to construct the "target-matrices" (2) for the SeDJoCo equations. NCG [6] was used to solve SeDJoCo with an initial guess $\boldsymbol{B}_0 = \boldsymbol{I}$ to obtain $\widehat{\boldsymbol{B}}_0$. Then, we used the proposed (approximate) ML "identification-correction" scheme to obtain the final estimate of the demixing matrix $\widehat{\boldsymbol{B}}$. Our results are based on 1000 independent trials.

Fig. 2 shows the average empirical total ISR (as defined, e.g., in [10]) versus T, obtained by the "non-corrected" ($\widehat{\boldsymbol{B}}_0$) and "corrected" ($\widehat{\boldsymbol{B}}$) solutions. This average is taken over the "best" (lowest ISR) 99% trials of $\widehat{\boldsymbol{B}}_0$ and $\widehat{\boldsymbol{B}}$ (separately), since in some (rare) trials $\widehat{\boldsymbol{B}}_0 \notin \mathcal{P}$, so our model assumption (4) does not hold, and the proposed scheme does not (necessarily) yield the ML solution. In such "outlier" cases, the solution $\widehat{\boldsymbol{B}}$ might be so "bad", in terms of ISR, that a single realization can devastate the empirical ISR average. As mentioned, we address this by discarding the worst 1% of the trials. However, it is still possible that $\widehat{\boldsymbol{B}}_0 \notin \mathcal{P}$, and $\widehat{\boldsymbol{B}}$ would be the ML solution, or that $\widehat{\boldsymbol{B}}_0 \in \mathcal{P}$, and $\widehat{\boldsymbol{B}}$ would not be the ML solution, since $\widehat{\boldsymbol{B}}$ does not depend directly on the distance of $\widehat{\boldsymbol{B}}_0$ from the ML solution, but rather on the stationary point to which the algorithm converges when initialized by

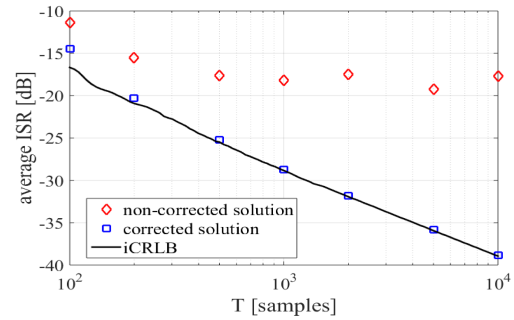

Fig. 2: Average ISR vs. T of the "non-corrected" and "corrected" SeDJoCo solutions, compared with the iCRLB (whose apparent non-smoothness is due to the different temporal variance-profiles of the sources). Results are based on 1000 independent trials, where the average was taken over the "best" 99% trials (for each solution, separately) as explained in the text.

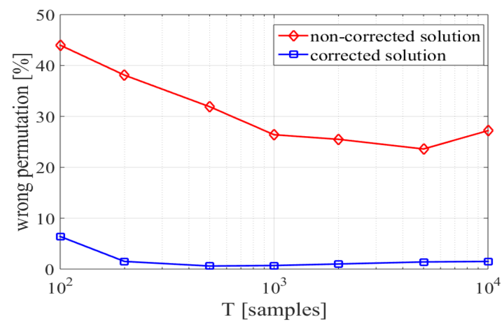

Fig. 3: Empirical percentage for a wrong permutation solution (and hence, not the ML solution) vs. T. Based on averaging 1000 independent trials.

$\widehat{\boldsymbol{B}}_0$. Considering this, the improvement in ISR between the non-corrected and corrected estimates is easily seen, with the corrected estimate attaining the iCRLB, implying optimality.

Additionally, we present in Fig. 3 the empirical probability (versus T) for a non-corrected / corrected solution to be "wrongly permuted", based on all 1000 trials (for each T). To this end, the wrong permutations were identified using our knowledge of the true mixing matrix in each trial. Evidently, our ML "identifiction-correction" scheme reduces the probability of a wrong permutation substantially, thus increasing the overall efficiency of the entire separation process.

5. CONCLUSION

In the context of the Gaussian semi-BSS problem, we proposed an approximate ML "identification-correction" scheme for solutions of the SeDJoCo problem which are not the ML solution. The proposed scheme is general and appropriate for any set of Gaussian sources (not necessarily stationary) with known temporal covariance matrices. In comparison to the naïve approach, by which any SeDJoCo solution is determined to be the estimated demixing matrix, our method shows significant improvement in the resulting empirical probability of obtaining the ML solution, hence leading to improved ISR (optimal in most cases).

6. REFERENCES

[1] D.-T. Pham, "Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion," *Signal Processing*, vol. 81, no. 4, pp. 855–870, 2001.

[2] M. Ghogho, A. Swami, and T. Durrani, "Approximate maximum likelihood blind source separation with arbitrary source PDFs," in *Statistical Signal and Array Processing (SSAP'2000)*, 2000, pp. 368–372.

[3] K. Todros and J. Tabrikian, "QML-based joint diagonalization of positive-definite Hermitian matrices," *IEEE Trans. on Signal Processing*, vol. 58, no. 9, pp. 4656–4673, 2010.

[4] A. 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.

[5] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.

[6] A. Yeredor, B. Song, F. Roemer, and M. Haardt, "A sequentially drilled joint congruence (SeDJoCo) transformation with applications in blind source separation and multiuser MIMO systems," *IEEE Trans. on Signal Processing*, vol. 60, no. 6, pp. 2744–2757, 2012.

[7] B. Song, F. Roemer, and M. Haardt, "Using a new structured joint congruence (STJoCo) transformation of Hermitian matrices for precoding in multi-user MIMO systems," in *ICASSP'2010*, 2010, pp. 3414–3417.

[8] S. Dégerine and A. 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.

[9] A. Yeredor, Y. Cheng, and M. Haardt, "On multiple solutions of the "sequentially drilled" joint congruence transformation (SeDJoCo) problem for semi-blind source separation," in *ICASSP'2016*. IEEE, 2016.

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