ON CONSISTENCY AND ASYMPTOTIC UNIQUENESS IN QUASI-MAXIMUM LIKELIHOOD BLIND SEPARATION OF TEMPORALLY-DIVERSE SOURCES

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

[†]School of Electrical Engineering
Tel-Aviv University
{amirwei2@mail, arie@eng}.tau.ac.il

[⋆]Communications Research Laboratory
Ilmenau University of Technology
{sher-ali.cheema, martin.haardt}@tu-ilmenau.de

ABSTRACT

In its basic, fully blind form, Independent Component Analysis (ICA) does not rely on a particular statistical model of the sources, but only on their mutual statistical independence, and therefore does not admit a Maximum Likelihood (ML) estimation framework. In semi-blind scenarios statistical models of the sources are available, enabling ML separation. Quasi-ML (QML) methods operate in the (more realistic) fully-blind scenarios, simply by presuming some hypothesized statistical models, thereby obtaining QML separation. When these models are (or are assumed to be) Gaussian with distinct temporal covariance matrices, the (quasi-)likelihood equations take the form of a "Sequentially Drilled Joint Congruence" (SeDJoCo) transformation problem. In this work we state some mild conditions on the sources' true and presumed covariance matrices, which guarantee consistency of the QML separation when the SeDJoCo solution is asymptotically unique. In addition, we derive a necessary "Mutual Diversity" condition on these matrices for the asymptotic uniqueness of the SeDJoCo solution. Finally, we demonstrate the consistency of QML in various simulation scenarios.

Index Terms— Quasi-maximum likelihood, blind source separation, consistency, SeDJoCo.

1. INTRODUCTION

Classically, the Independent Component Analysis (ICA) framework relies only on statistical independence of the sources, and does not employ any further statistical model assumptions. This strong, robust and often well-justified paradigm gave rise to some classical, well-known model-free separation approaches, such as mutual information minimization (e.g., [1, 2]), high order moments based methods (e.g., [3, 4]) and approximate joint diagonalization based on Second-Order Statistics (SOS) (e.g., [5–7]), to name a few.

In the context of "semi-blind" separation (e.g., [8, 9]), some statistical information is assumed to be known *a-priori*, enabling to obtain the Maximum Likelihood (ML) estimate of the mixing matrix. In particular, for temporally-correlated Gaussian sources with known (distinct) temporal covariance structures, the resulting likelihood equations take a special form of joint matrix transformations (reminiscent of, but essentially different from, approximate joint diagonalization), termed a "Sequentially Drilled" Joint Congruence (SeDJoCo) transformation in [10] (see also [11] and [12]).

Quasi Maximum Likelihood (QML) approaches (e.g., [13–16]), on the other hand, make some model assumptions on the sources, and use some "educated guess" for the associated parameters, in order to facilitate a "quasi-" ML estimate, which would hopefully approximate the ML estimate when the assumed model is close to reality. Pham and Garat proposed in [17] two QML methods, one of which is tailored to temporally-correlated stationary sources with distinct spectra. Presuming that the sources are Gaussian with known spectra, the implied likelihood of the observed mixtures is expressed and maximized (with respect to the unknown mixing matrix), essentially resulting in a set of SeDJoCo equations. However, since the sources are not necessarily Gaussian, and their spectra are actually unknown, in this case the resulting SeDJoCo equations are in fact the *quasi*-likelihood (rather than the likelihood) equations.

While the true ML estimate enjoys some appealing, well-known properties, such as consistency and asymptotic efficiency [18], these properties are generally not shared by QML estimates. Our goal in this paper is to outline some mild conditions (not only on the signals, but also on their presumed covariance matrices) for the asymptotic uniqueness of the SeDJoCo solution and the resulting consistency[1] of the QML estimate.

2. PROBLEM FORMULATION

Consider the classical linear mixture model

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

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 signals,

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.

[1]An estimate is considered consistent (with the inevitable scale and permutation ambiguities) in the context of ICA if its resulting interference-to-source ratios (ISRs) all tend to zero (i.e., perfect separation) when the observation lengths tend to infinity.

each of length T, and $\boldsymbol{X}$ is the observed mixtures matrix, from which it is desired to estimate the demixing-matrix $\boldsymbol{B} \triangleq \boldsymbol{A}^{-1}$ in order to separate (estimate) the source signals. We start by briefly reviewing the semi-blind scenario with Gaussian sources and the corresponding ML approach, which leads to the SeDJoCo equations. We then move to the fully-blind scenario and formulate the QML approach, in which the true (but unknown) statistical model is replaced by a presumed model.

2.1. The ML Approach for Gaussian Semi-BSS

As shown in [9, 13] and [19] (chapter 7), when the source signals are zero-mean Gaussian, each with a known, Positive-Definite (PD) 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 $\widehat{\boldsymbol{B}}_{\mathrm{ML}}$ is a solution of the following set of nonlinear equations:

$$\widehat{\boldsymbol{B}}\boldsymbol{Q}_k\widehat{\boldsymbol{B}}^{\mathrm{T}}\boldsymbol{e}_k = \boldsymbol{e}_k, \quad \forall k \in \{1,\ldots,K\}, \tag{2}$$

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

$$\boldsymbol{Q}_k \triangleq \frac{1}{T}\boldsymbol{X}\boldsymbol{C}_k^{-1}\boldsymbol{X}^{\mathrm{T}} \in \mathbb{R}^{K\times K}, \quad \forall k \in \{1,\ldots,K\}, \tag{3}$$

are termed the "target-matrices". The solution $\widehat{\boldsymbol{B}}$ of (2) jointly transforms the set of target-matrices so that the k-th column (and, by symmetry of $\boldsymbol{Q}_k$, also the k-th row) of the k-th transformed matrix $\widehat{\boldsymbol{B}}\boldsymbol{Q}_k\widehat{\boldsymbol{B}}^{\mathrm{T}}$ equals the vector $\boldsymbol{e}_k$.

As shown in [10], a solution $\widehat{\boldsymbol{B}}$ always exists, but is (generally) not unique, as $K!-1$ other essentially different (for $K>2$) solutions may be characterized as in [20,21]. The solutions of (2) which are not the global maximizer are merely stationary points of the likelihood.

2.2. The Gaussian QML Approach for BSS

For the model (1), in a fully blind scenario no prior knowledge on the sources is available (except for their mutual statistical independence), neither in terms of their full distributions, nor in terms of any other statistical property, such as their temporal correlations. Thus, in a QML approach, a given hypothetical model of sources is assumed (incorporating their statistical independence), which is hopefully (but not necessarily) "close" to reality. The estimated demixing matrix $\widehat{\boldsymbol{B}}_{\mathrm{QML}}$ is then obtained as a solution of the likelihood equations for the assumed hypothetical model, which are commonly termed the "quasi-likelihood equations".

One plausible QML approach is to assume that all sources are zero-mean Gaussian with some (distinct) PD temporal covariance matrices $\widetilde{\boldsymbol{C}}_k$, thereby obtaining the SeDJoCo equations (2) once again, only now they are quasi-likelihood equations, and the target-matrices (3) are simply redefined as

$$\boldsymbol{Q}_k \triangleq \frac{1}{T}\boldsymbol{X}\boldsymbol{P}_k\boldsymbol{X}^{\mathrm{T}}, \quad \forall k \in \{1,\ldots,K\}, \tag{4}$$

where $\boldsymbol{P}_k = \widetilde{\boldsymbol{C}}_k^{-1}$. Of course, these $\boldsymbol{P}_k$ matrices are not necessarily (and in practice almost never are) the true inverse covariance matrices. Nonetheless, when they are chosen appropriately, according to mild conditions specified in the next section, the resulting SeDJoCo solution (or at least one of the solutions, if the solution is not unique) is guaranteed to be a consistent estimate.

Note that if the sources are indeed Gaussian and each $\boldsymbol{P}_k$ somehow happens to coincide with $\boldsymbol{C}_k^{-1}$, the QML estimate essentially becomes the ML estimate. We shall therefore refer to a QML estimate as a "(Q)ML" estimate whenever referring to a property shared by both the QML and ML estimates.

3. CONDITIONS FOR CONSISTENCY & ASYMPTOTIC UNIQUENESS OF (Q)ML

We begin by noting that for $\boldsymbol{A}=\boldsymbol{I}$ (a "non-mixing" condition), the (Q)ML target-matrices $\boldsymbol{Q}_k$ (4) are *asymptotically diagonal* under some mild conditions stated in the Lemma below. This important property will be used later to establish consistency of the resulting estimates under mixing conditions.

Lemma 1 (Asymptotic diagonality of the target-matrices when $\boldsymbol{X}=\boldsymbol{S}$) *Let us temporarily denote the observation-length-dependent (PD) covariance matrices and their presumed inverses (resp.) as $\boldsymbol{C}_k^{(T)}, \boldsymbol{P}_k^{(T)} \in \mathbb{R}^{T\times T}$ for an observation length T. Consider the conditions:*

1. *The following limits exist and are finite and positive:*

   $$\phi_{k,\ell} \triangleq \lim_{T\to\infty}\tfrac{1}{T}\mathrm{Tr}\left(\boldsymbol{P}_k^{(T)}\boldsymbol{C}_\ell^{(T)}\right) > 0 \quad \forall k,\ell \in \{1,\ldots,K\}.$$

2. *All matrices $\boldsymbol{C}_k^{(T)}$, $\boldsymbol{P}_k^{(T)}$ are element-wise bounded by an exponentially-decaying Toeplitz matrix, namely there exist some finite ρ and a positive α, such that*

   $$|C_k^{(T)}[m,n]|, |P_k^{(T)}[m,n]| < \rho^2\cdot e^{-\alpha|m-n|}$$
   $$\forall k \in \{1,\ldots,K\}, \forall m,n \in \{1,\ldots,T\}, \forall T \in \mathbb{Z}^+$$

3. *The 4-th order joint cumulants of each source are similarly bounded by an exponentially decaying function of the time differences, i.e., there exist some finite ϱ and a positive β, such that for each $k \in \{1,\ldots,K\}$*

   $$|\mathrm{cum}(s_k[m],s_k[n],s_k[p],s_k[q])|$$
   $$< \varrho^4\cdot e^{-\beta(|m-n|+|p-q|+|m-p|+|n-q|+|m-q|+|n-p|)}$$

   for all m,n,p,q, where $\mathrm{cum}(\cdot,\cdot,\cdot,\cdot)$ denotes the 4-th order joint cumulant of its arguments.

Under these conditions the following property holds:
For $\boldsymbol{A}=\boldsymbol{I}$ (so that $\boldsymbol{X}=\boldsymbol{S}$), the (Q)ML target-matrices $\boldsymbol{Q}_k$ are asymptotically diagonal for all $k \in \{1,\ldots,K\}$,

$$\boldsymbol{A}=\boldsymbol{I}: \quad \boldsymbol{Q}_k = \frac{1}{T}\boldsymbol{S}\boldsymbol{P}_k^{(T)}\boldsymbol{S}^{\mathrm{T}} \xrightarrow[T\to\infty]{m.s.} \boldsymbol{\Phi}_k, \tag{5}$$

where $\mathbf{\Phi}_k \triangleq \mathrm{Diag}(\phi_{k,1}, \phi_{k,2}, \ldots, \phi_{k,K})$ *is a PD diagonal matrix with* $\phi_{k,1}, \ldots, \phi_{k,K}$ *(defined in Condition 1) along its diagonal, and where the convergence is in the mean square sense [22].*

Note that the conditions of the Lemma are quite loose and are readily satisfied, e.g., whenever the sources are stationary parametric linear processes (Auto-Regressive (AR), Moving-Average (MA) or ARMA) with finite 4-th order moments and the $\boldsymbol{P}_k^{(T)}$ matrices are symmetric banded Toeplitz matrices, but also for many other types of non-stationary signals and non-Toeplitz matrices, as long as the $\boldsymbol{C}_k^{(T)}$ and $\boldsymbol{P}_k^{(T)}$ matrices, as well as the joint cumulants tensor, all have bounded diagonals and a sufficient rate of decay of their elements away from their diagonals.

Due to space limitations, a rigorous proof is omitted from here; Note only that the mean of $\boldsymbol{Q}_k[p,q] = \frac{1}{T}\boldsymbol{s}_p^{\mathrm{T}}\boldsymbol{P}_k^{(T)}\boldsymbol{s}_q$ reads

$$E\left[\boldsymbol{Q}_k[p,q]\right] = \frac{1}{T}\mathrm{Tr}\left(\boldsymbol{P}_k^{(T)} E[\boldsymbol{s}_q\boldsymbol{s}_p^{\mathrm{T}}]\right) = \begin{cases} \frac{1}{T}\mathrm{Tr}\left(\boldsymbol{P}_k^{(T)}\boldsymbol{C}_p^{(T)}]\right) \xrightarrow[T\to\infty]{} \phi_{k,\ell} & p=q \\ 0 & p \neq q \end{cases}, \quad (6)$$

and using Conditions 2 and 3 it can be shown that their variances tend to zero (note further that the exponential decay of the bound in these conditions is sufficient but not necessary).

This important property of asymptotic diagonality would guarantee the consistency of an estimate obtained as a SeDJoCo solution, as long as this solution is asymptotically unique. While the conditions of Lemma 1 can be easily satisfied by the $\boldsymbol{P}_k$ matrices, there is an additional necessary condition that has to be satisfied by these matrices (together with the true $\boldsymbol{C}_k$ matrices) in order to obtain such asymptotic uniqueness. To derive this condition we note the following.

Let $\widehat{\boldsymbol{B}}_{\mathrm{QML}}$ denote a (specific) solution of the SeDJoCo equations (2) with target-matrices defined as in (4). By Lemma 1 we have, for all $k \in \{1, \ldots, K\}$,

$$\begin{aligned}
&\widehat{\boldsymbol{B}}_{\mathrm{QML}}\boldsymbol{Q}_k\widehat{\boldsymbol{B}}_{\mathrm{QML}}^{\mathrm{T}} = \widehat{\boldsymbol{B}}_{\mathrm{QML}}\left(\frac{1}{T}\boldsymbol{X}\boldsymbol{P}_k\boldsymbol{X}^{\mathrm{T}}\right)\widehat{\boldsymbol{B}}_{\mathrm{QML}}^{\mathrm{T}} = \\
&\widehat{\boldsymbol{B}}_{\mathrm{QML}}\left(\frac{1}{T}\boldsymbol{A}\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{S}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\right)\widehat{\boldsymbol{B}}_{\mathrm{QML}}^{\mathrm{T}} = \\
&\left(\widehat{\boldsymbol{B}}_{\mathrm{QML}}\boldsymbol{A}\right)\left(\frac{1}{T}\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{S}^{\mathrm{T}}\right)\left(\widehat{\boldsymbol{B}}_{\mathrm{QML}}\boldsymbol{A}\right)^{\mathrm{T}} \triangleq \\
&\widehat{\boldsymbol{G}}\left(\frac{1}{T}\boldsymbol{S}\boldsymbol{P}_k\boldsymbol{S}^{\mathrm{T}}\right)\widehat{\boldsymbol{G}}^{\mathrm{T}} \xrightarrow[T\to\infty]{\text{Lemma 1}} \widehat{\boldsymbol{G}}\mathbf{\Phi}_k\widehat{\boldsymbol{G}}^{\mathrm{T}},
\end{aligned}$$

where we have defined $\widehat{\boldsymbol{G}} \triangleq \widehat{\boldsymbol{B}}_{\mathrm{QML}}\boldsymbol{A}$ as the QML estimated global demixing-mixing matrix. Thus, since $\widehat{\boldsymbol{B}}_{\mathrm{QML}}$ is a solution of (2), the implied asymptotic SeDJoCo equations, expressed in terms of $\widehat{\boldsymbol{G}}$, take the form

$$\widehat{\boldsymbol{G}}\mathbf{\Phi}_k\widehat{\boldsymbol{G}}^{\mathrm{T}}\boldsymbol{e}_k = \boldsymbol{e}_k, \quad \forall k \in \{1, \ldots, K\}. \quad (7)$$

It is easily seen that the diagonal matrix (which implies perfect separation)

$$\widehat{\boldsymbol{G}}_{\mathrm{o}} = \mathrm{Diag}\left(\phi_{1,1}^{-1/2}, \phi_{2,2}^{-1/2}, \cdots, \phi_{K,K}^{-1/2}\right) \quad (8)$$

solves (7) , and consequently we have

$$\widehat{\boldsymbol{G}}_{\mathrm{o}}\mathbf{\Phi}_k\widehat{\boldsymbol{G}}_{\mathrm{o}}^{\mathrm{T}} = \mathrm{Diag}\left(\frac{\phi_{k,1}}{\phi_{1,1}}, \frac{\phi_{k,2}}{\phi_{2,2}}, \ldots, \frac{\phi_{k,K}}{\phi_{K,K}}\right). \quad (9)$$

However, the following is a necessary condition for the uniqueness of $\widehat{\boldsymbol{G}}_{\mathrm{o}}$ as a solution of (7) (and therefore also of the respective separating solution $\widehat{\boldsymbol{B}}_{\mathrm{o}} \triangleq \widehat{\boldsymbol{G}}_{\mathrm{o}}\boldsymbol{A}^{-1}$ as a solution of (2)).
<u>The “Mutual Diversity” condition:</u>

$$\begin{aligned} \forall i \neq j \in \{1, \ldots, K\}: \quad \exists k \neq \ell \in \{1, \ldots, K\}: \\ \phi_{k,i} \cdot \phi_{\ell,j} \neq \phi_{\ell,i} \cdot \phi_{k,j}. \end{aligned} \quad (10)$$

To show its necessity, assume that it is not satisfied, so that

$$\begin{aligned} &\exists i \neq j \in \{1, \ldots, K\}: \quad \forall k \neq \ell \in \{1, \ldots, K\}: \\ &\phi_{k,i} \cdot \phi_{\ell,j} = \phi_{\ell,i} \cdot \phi_{k,j} \quad \Rightarrow \quad \frac{\phi_{k,i}}{\phi_{k,j}} = \frac{\phi_{\ell,i}}{\phi_{\ell,j}} \triangleq \mu. \end{aligned} \quad (11)$$

Now define the diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{K\times K}$ in which $\Lambda[j,j] = \sqrt{\mu\frac{\phi_{j,j}}{\phi_{i,i}}}$ and $\Lambda[k,k] = 1 \quad \forall k \neq i$. Then

$$\mathbf{\Lambda}\widehat{\boldsymbol{G}}_{\mathrm{o}}\mathbf{\Phi}_k\widehat{\boldsymbol{G}}_{\mathrm{o}}^{\mathrm{T}}\mathbf{\Lambda}^{\mathrm{T}} = \mathrm{Diag}\left(\frac{\phi_{k,1}}{\phi_{1,1}}, \ldots, \frac{\phi_{k,i}}{\phi_{i,i}}, \ldots, \frac{\phi_{k,i}}{\phi_{i,i}}, \ldots, \frac{\phi_{k,K}}{\phi_{K,K}}\right), \quad (12)$$

where the two equal elements are in its $[i,i]$-th and $[j,j]$-th positions. Therefore, if we now define a rotation matrix $\boldsymbol{U} \in \mathbb{R}^{K\times K}$ along the i and j coordinates, such that

$$U[k,\ell] = \begin{cases} 1 & k=\ell\neq i,j \\ \cos(\theta) & k=\ell=i, k=\ell=j \\ \sin(\theta) & k=i, \ell=j \\ -\sin(\theta) & k=j,\ell=i \\ 0 & \text{otherwise} \end{cases}, \quad (13)$$

and θ is some arbitrary rotation angle, multiplying by $\boldsymbol{U}$ on the right and by $\boldsymbol{U}^{\mathrm{T}}$ on the left will have no further effect:

$$\boldsymbol{U}\mathbf{\Lambda}\widehat{\boldsymbol{G}}_{\mathrm{o}}\mathbf{\Phi}_k\widehat{\boldsymbol{G}}_{\mathrm{o}}^{\mathrm{T}}\mathbf{\Lambda}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} = \mathbf{\Lambda}\widehat{\boldsymbol{G}}_{\mathrm{o}}\mathbf{\Phi}_k\widehat{\boldsymbol{G}}_{\mathrm{o}}^{\mathrm{T}}\mathbf{\Lambda}^{\mathrm{T}}. \quad (14)$$

If we now multiply by $\mathbf{\Lambda}^{-1}$ on the left and on the right, we would undo the operation in (12) and satisfy the SeDJoCo equations again (as in (9)), that is

$$\mathbf{\Lambda}^{-1}\boldsymbol{U}\mathbf{\Lambda}\widehat{\boldsymbol{G}}_{\mathrm{o}}\mathbf{\Phi}_k\widehat{\boldsymbol{G}}_{\mathrm{o}}^{\mathrm{T}}\mathbf{\Lambda}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\mathbf{\Lambda}^{-\mathrm{T}} = \mathrm{Diag}\left(\frac{\phi_{k,1}}{\phi_{1,1}}, \frac{\phi_{k,2}}{\phi_{2,2}}, \ldots, \frac{\phi_{k,K}}{\phi_{K,K}}\right), \quad (15)$$

so that the matrix $\widehat{\boldsymbol{G}}_* \triangleq \boldsymbol{\Lambda}^{-1}\boldsymbol{U}\boldsymbol{\Lambda}\widehat{\boldsymbol{G}}_{\rm o}$, which is obviously not diagonal (for $\theta \neq 0, \pi$), also solves (7), and therefore in this case the separating solution $\widehat{\boldsymbol{B}}_{\rm o}$ is not a unique solution of the SeDJoCo equations (2) (not even asymptotically).

The "Mutual Diversity" (MD) condition essentially determines mutual limitations on the hypothesized $\boldsymbol{P}_k$ and true $\boldsymbol{C}_k$ matrices. Note first, that the basic ICA SOS identifiability condition (e.g., [9]), which requires that all $\boldsymbol{C}_k$ matrices be distinct (i.e., no two matrices in the set are identical up to multiplication by a constant) is already contained in the MD condition (as could be expected), since if, say, $\boldsymbol{C}_i = \gamma \cdot \boldsymbol{C}_j$ (for some $i \neq j \in \{1, \dots, K\}$ and some constant γ), then for these i, j, regardless of the choice of $\boldsymbol{P}_k$ matrices, we shall always have $\phi_{k,i} \cdot \phi_{\ell,j} = \phi_{\ell,i} \cdot \phi_{k,j}$ ($\forall k, \ell \in \{1, \dots, K\}$), and the MD condition would be breached. Therefore, the MD condition generalizes the ICA SOS identifiability condition.

Moreover, in the (semi-blind) case of ML estimation, when all $\boldsymbol{P}_k$ are the true $\boldsymbol{C}_k^{-1}$ (resp.), we have $\phi_{k,k} = 1$ for all k, and (if all $\boldsymbol{C}_k$ are distinct) also $\phi_{k,\ell} \cdot \phi_{\ell,k} > 1$ for all $k \neq \ell$ (by applying Jensen's inequality [23] with the convex function $\phi(x) = 1/x, x \in (0, \infty)$ to the eigenvalues of $\boldsymbol{C}_k^{-1}\boldsymbol{C}_\ell$), and therefore the MD condition is automatically satisfied for all $i \neq j$ with $k = i$, $\ell = j$.

However, in the case of QML estimation, the choice of $\boldsymbol{P}_k$ matrices must consider the MD condition. Note, for example, that the condition forbids the use of the same matrix $\boldsymbol{P}$ (or multiples thereof by a constant) for all $\boldsymbol{P}_k$ (but generally does not forbid the use of two or more identical matrices in the set).

Although we have only shown that the MD condition is *necessary* for the asymptotic uniqueness of the SeDJoCo solution, according to our experience it is usually also sufficient (although counter-examples do exist). In the following section we demonstrate the consistency of the resulting QML estimate in various scenarios.

4. SIMULATION RESULTS

We simulated three different scenarios, all sharing the following setup. A set of $K = 4$ sources, all MA processes of order 5 (MA(5)), was generated by filtering statistically independent, zero-mean, unit-variance, temporally-i.i.d. noises, denoted $\{w_k[n]\}_{k=1}^4$, using 4 different fixed Finite Impulse Response (FIR) filters of length 5 with randomly drawn coefficients. 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 (4) for the SeDJoCo equations. Newton's Conjugate Gradient [10] was used to solve SeDJoCo with an initial guess $\widehat{\boldsymbol{B}} = \boldsymbol{I}$, followed by the "identification-correction" scheme [21]. Our results are based on 10^4 independent trials.

In the first scenario $\{w_k[n]\}_{k=1}^4$ were all Gaussian (hence so are the sources), but the $\{\boldsymbol{P}_k\}_{k=1}^4$ matrices were taken as the inverses of correlation matrices of arbitrary MA(4) processes (unrelated to the true sources). In the second scenario, all $\boldsymbol{P}_k$ matrices were taken as the inverses of the true $\boldsymbol{C}_k$ matrices (resp.), but $\{w_k[n]\}_{k=1}^4$ were all *Uniformly* distributed. In the third scenario, $\{w_k[n]\}_{k=1}^4$ were all *Laplace* distributed, and (similarly to the first experiment) the $\boldsymbol{P}_k$ matrices were taken as the inverses of the correlation matrices of MA(3) processes. Obviously, in all three scenarios the resulting SeDJoCo equations are "quasi"-likelihood equations (either due to "wrong" $\boldsymbol{P}_k$ matrices, or to a non-Gaussian sources distribution, or both). All the conditions of Lemma 1 were naturally satisfied by construction, and we verified that the MD condition was satisfied as well.

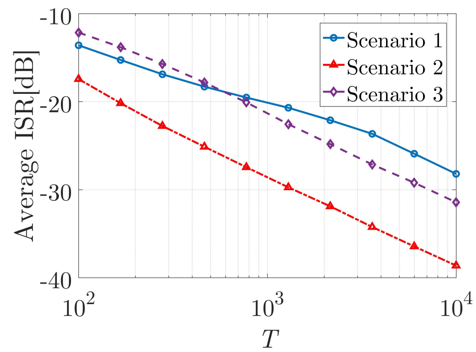

Fig. 1: Average ISR vs. T for the three scenarios. Scenario 1: Gaussian driving-noises with arbitrary $\boldsymbol{P}_k$'s, Scenario 2: Uniform driving-noises with $\boldsymbol{P}_k = \boldsymbol{C}_k^{-1}$, Scenario 3: Laplace driving-noises with arbitrary $\boldsymbol{P}_k$'s. Results are based on 10^4 independent trials

Fig. 1 shows the average empirical total ISR (as defined, e.g., in [9]) versus T, obtained by solving the QML SeDJoCo equations (2) as described above. Clearly, the ISR decreases monotonically and a consistency trend is evident. We remind that in the first and third scenarios we chose the $\boldsymbol{P}_k$ matrices arbitrarily (and randomly), in order to demonstrate the robustness of the QML estimate. In practice, however, an "educated guess" for the sources' covariance matrices may be available, which would directly affect (and improve) the resulting ISR. In fact, it is our purpose to extend this work to a more general performance analysis of the QML estimate in future work.

We note in passing that although the sources considered in these simulations are stationary, our theoretical results are valid for any set of sources (and $\boldsymbol{P}_k$ matrices) that satisfies the conditions of Lemma 1 and the MD condition, which clearly do not require stationarity.

5. CONCLUSION

In the context of blind separation of temporally-diverse sources using the Gaussian QML approach, we presented some mild conditions on the sources' covariance and cumulants structures (trivially satisfied for a wide range of sources), as well as on the hypothesized covariance matrices, under which the asymptotic diagonality of the "target-matrices" for the quasi-likelihood (SeDJoCo) equations is guaranteed in a non-mixing condition. Such diagonality guarantees, in turn, the consistency of at least one of the SeDJoCo solutions. In addition, we derived the MD condition, which is necessary for the asymptotic uniqueness of that solution, and generalizes the ICA SOS identifiability condition.

6. REFERENCES

[1] H. H. Yang and S. Amari, "Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information," *Neural computation*, vol. 9, no. 7, pp. 1457–1482, 1997.

[2] D. T. Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1935–1946, 2002.

[3] J.-F. Cardoso, "Source separation using higher order moments," in *Proc., IEEE ICASSP*, 1989, pp. 2109–2112.

[4] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural computation*, vol. 11, no. 1, pp. 157–192, 1999.

[5] A. Belouchrani and K. Abed-Meraim, "Séparation aveugle au second ordre de sources corrélées," in *14 Colloque sur le traitement du signal et des images, FRA, 1993*. GRETSI, Groupe dEtudes du Traitement du Signal et des Images, 1993.

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

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

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

[9] 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.

[10] 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.

[11] A. Yeredor, "On hybrid exact-approximate joint diagonalization," in *In Proc., 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2009, pp. 312–315.

[12] 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 *Proc., IEEE ICASSP*, 2010, pp. 3414–3417.

[13] 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.

[14] 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.

[15] 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.

[16] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky, "Quasi maximum likelihood MIMO blind deconvolution: Super-and sub-Gaussianity versus consistency," *IEEE Trans. on Signal Processing*, vol. 53, no. 7, pp. 2576–2579, 2005.

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

[18] H. Van Trees, K. Bell, and Z. Tian, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. Wiley, 2013.

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

[20] 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 *Proc. ICASSP*, 2016, pp. 2916–2920.

[21] A. Weiss, A. Yeredor, S. A. Cheema, and M. Haardt, "A maximum likelihood "identification-correction" scheme of sub-optimal "SeDJoCo" solutions for semi-blind source separation," in *Proc. ICASSP*, 2017, pp. 4158–4162.

[22] A. F. Karr, "Convergence of random variables," in *Probability*. Springer, 1993, pp. 135–162.

[23] J. L. W. V. Jensen, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta mathematica*, vol. 30, no. 1, pp. 175–193, 1906.