Online Adaptive Quasi-Maximum Likelihood Blind Source Separation of Stationary Sources

Amir Weiss and Arie Yeredor
School of Electrical Engineering, Tel-Aviv University, Israel
amirwei2@mail.tau.ac.il, arie@eng.tau.ac.il

***Abstract*—In the context of Blind Source Separation (BSS), we consider the problem of online separation of stationary sources. Based on the Maximum Likelihood (ML) solution for semi-blind separation of temporally-diverse Gaussian sources, and assuming that a parametric model of the sources' spectra is available, we propose an online adaptive Quasi-ML (QML) separation algorithm. The algorithm operates in an alternating fashion, updating at each iteration the (nuisance) spectra-characterizing parameters first, and then the demixing-matrix estimates, according to simple, computationally efficient update expressions which we derive. Our proposed algorithm, which leads to consistent separation of the sources, is demonstrated here, both analytically and empirically in a simulation experiment, for first-order autoregressive sources.**

***Index Terms*—Blind source separation, independent component analysis, quasi-maximum likelihood, SeDJoCo.**

I. Introduction

In its most general form, Independent Component Analysis (ICA) 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 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 (AJD) 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. One particularly interesting case is for (mutually independent) temporally-correlated Gaussian sources with known (distinct) temporal covariance structures.

More formally, 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, $\boldsymbol{S} = \left[\boldsymbol{s}_1 \; \boldsymbol{s}_2 \; \cdots \; \boldsymbol{s}_K\right]^{\mathrm{T}} \in \mathbb{R}^{K\times T}$ is an unknown sources' matrix of K statistically independent, zero-mean Gaussian processes, each of length T 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[1], and $\boldsymbol{X}$ is the observed mixtures matrix (where its k-th row is the received signal at the k-th sensor), from which it is desired to estimate the demixing-matrix $\boldsymbol{B} \triangleq \boldsymbol{A}^{-1}$ in order to separate (estimate) the source signals. As shown in [9], [10] and [11] (chapter 7), the ML estimate $\widehat{\boldsymbol{B}}_{\mathrm{ML}}$ in this case is a solution of the following set of nonlinear equations:

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

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

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

are termed the "target-matrices". The resulting likelihood equations (2) are termed a "Sequentially Drilled" Joint Congruence (SeDJoCo) transformation in [12] (see also [13] and [14]), since the solution $\widehat{\boldsymbol{B}}$ of (2) jointly transforms the (ordered) set of target-matrices so that the k-th column (and, by symmetry of $\boldsymbol{Q}_k^{(T)}$, also the k-th row) of the k-th transformed matrix $\widehat{\boldsymbol{B}}\boldsymbol{Q}_k^{(T)}\widehat{\boldsymbol{B}}^{\mathrm{T}}$ equals the vector $\boldsymbol{e}_k$.

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 Quasi Maximum Likelihood (QML) approach, a given hypothetical model of sources is assumed, 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, presented by Pham and Garat in [15], is designed for temporally-correlated stationary sources with distinct spectra. Presuming that the sources are Gaussian with some hypothesized spectra, the implied likelihood of the observed mixtures is expressed and maximized (with respect to (w.r.t.) 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, the resulting SeDJoCo equations are in fact the *quasi*-likelihood (rather than the likelihood) equations.

In this work, we address two aspects of the Gaussian QML approach, which, to the best of our knowledge, were not addressed in the literature so far. The first is regarding the sources' spectra presumption, which is likely to be (at least) inaccurate in a fully blind scenario. Recall that the demixing matrix' QML Estimate (QMLE) is a SeDJoCo solution for the set of target-matrices (3), constructed using the measurements

[1] In the sense that none of these covariance matrices is merely a scaled version of another.

978-1-5386-6378-3/18/$31.00 ©2018 IEEE

matrix $\boldsymbol{X}$ and the presumed inverse temporal covariance matrices $\{\boldsymbol{C}_k^{-1}\}_{k=1}^K$. Naturally, the modeling error introduced by inaccuracies in the presumed $\{\boldsymbol{C}_k^{-1}\}_{k=1}^K$ would be translated into degraded separation performance characterized by a higher Interference-to-Source-Ratio (ISR), a common measure of the separation quality (e.g., [16]). Moreover, it was recently shown in [17] that the resulting ISR attained by the SeDJoCo QML solution is asymptotically independent of the sources' distributions, and is asymptotically dependent *only* on the sources' SOS. Thus, adaptation of the presumed initial spectra "towards" the true spectra of the sources (based on the observed mixtures), and a subsequent update of the SeDJoCo solution accordingly, would enable to attain lower ISRs.

The second aspect has to do with the mode of operation. Notice that in order to solve SeDJoCo and obtain $\widehat{\boldsymbol{B}}_{\text{QML}}$, one has to wait and accumulate all the available samples so as to account for all the available information. This restriction is not limiting when working in an offline mode and when the response time is not an important factor. However, when working in an online mode, i.e., when new samples of the mixtures arrive at every sampling period, and the current updated QMLE is required at that particular time instance, the offline "batch" approach is computationally wasteful due to its inability to efficiently use the precalculated target-matrices and the estimate already obtained at the previous sampling period. Hence, an online algorithm which updates the previous estimate using some small, efficiently computed update step is clearly attractive.

Motivated by these two aspects, we propose an adaptive online QML algorithm which takes them both into account. Assuming stationarity of the sources and some (fixed size) parametric model of their temporal covariance matrices, upon the arrival of each new vector-sample of mixtures, the proposed algorithm updates the presumed covariances, the target-matrices and accordingly the QMLE of the demixing matrix. This adaptive online scheme leads to constantly improving separation (on the average) due to the increasing sample size and the decreasing model-error inaccuracy.

The rest of this paper is organized as follows. In Section II we present the (blind) ICA problem and discuss the Gaussian QMLE. The online adaptive QML algorithm is presented in Section III, supported by simulations results presented in Section IV. Concluding remarks are given in Section V.

II. Problem Formulation

Consider model (1), only now the sources $\boldsymbol{S} \in \mathbb{R}^{K\times T}$ are not necessarily Gaussian, and their true covariance matrices $\boldsymbol{C}_k \in \mathbb{R}^{T\times T}$ are not necessarily known. Again, $\boldsymbol{X}$ is the observed mixtures matrix, from which it is desired to estimate the demixing-matrix $\boldsymbol{B}$ in order to separate the sources. Here, aside from the sources' mutual statistical independence, we also assume that the sources are stationary with distinct spectra, and that the spectrum of each source may be described by a known parametric model with a fixed number of deterministic unknown parameters. Therefore, given the parametric model of the sources' spectra, an "extended" Gaussian QML approach may be taken. In this approach (unlike in the "standard" Gaussian QML [17]), the unknown spectra-characterizing parameters are considered as nuisance parameters, whose consistent estimation would enable consequent consistent estimation of the demixing-matrix[2], which, in turn, implies consistent separation of the sources. Of course, theoretically, joint QML estimation of all the unknown parameters is possible by solving the resulting Gaussian quasi-likelihood equations for the demixing-matrix elements together with the spectra-characterizing parameters of all the sources. However, in this work we propose a somewhat simpler approach, which enables faster computations, and is therefore more suitable for online (possibly adaptive) separation.

III. The Online Adaptive Gaussian QML Algorithm

First, we note two key observations, on which the proposed solution scheme is based:

1) Given an estimate of the sources' spectra, the demixing matrix may be estimated by solving the corresponding SeDJoCo equations (2), where the target-matrices (3) are constructed using $\boldsymbol{X}$ and the sources' spectra estimates.
2) Given an estimate $\widehat{\boldsymbol{B}}$, the sources' spectra-characterizing parameters (and therefore the sources' spectra) may be estimated by QML estimation, presuming the estimated sources $\widehat{\boldsymbol{S}} \triangleq \widehat{\boldsymbol{B}}\boldsymbol{X}$ are (approximately) Gaussian.

Considering the above, and assuming some initial estimates of the sources' spectra-characterizing parameters are available, we propose to alternate between QML demixing-matrix estimation based on fixed sources' spectra estimates and QML sources' spectra-characterizing parameters estimation based on a fixed demixing-matrix estimate. Notice that this principle may be applied not only for online estimation, but also when the sample size is fixed.

Next, let us assume that for a sample size t, estimates of all the parameters involved are available, possibly as the results of a previous iteration. More specifically, denote $\{\widehat{\boldsymbol{C}}_k^{(t)}\}_{k=1}^K$ and $\widehat{\boldsymbol{B}}^{(t)}$ as the temporal covariance matrices' estimates (obtained from the estimates of spectra-characterizing parameters), and the demixing-matrix' estimate (obtained as a SeDJoCo solution), respectively, based on t samples from each sensor. For convenience, we now introduce the notation $\boldsymbol{X}^{(T)} \in \mathbb{R}^{K\times T}$ for the mixtures matrix representing a sample size T, as defined in (1). Upon the arrival of a new measurements vector $\boldsymbol{x}[t+1] = \boldsymbol{A}\boldsymbol{s}[t+1]$, where $\boldsymbol{s}[t+1] \in \mathbb{R}^{K\times 1}$ is the sources vector-sample at time instance $t+1$, new, more refined estimates of both $\{\boldsymbol{C}_k\}_{k=1}^K$ and $\boldsymbol{B}$ may be attained. The estimates of the sources' spectra-characterizing parameters may be updated using $\widehat{\boldsymbol{s}}^{(t)}[t+1] \triangleq \widehat{\boldsymbol{B}}^{(t)}\boldsymbol{x}[t+1]$, where $\widehat{\boldsymbol{s}}^{(t)}[t+1]$ denotes the sources' estimate at time instance $t+1$ based on the demixing-matrix estimate $\widehat{\boldsymbol{B}}^{(t)}$. Then, the demixing-matrix estimate is updated by a new SeDJoCo solution, based on $\boldsymbol{X}^{(t+1)}$ and the updated estimates $\{\widehat{\boldsymbol{C}}_k^{(t+1)}\}_{k=1}^K$.

[2] up to inevitable sign and scale ambiguities

To summarize, given the initial estimates $\widehat{\boldsymbol{B}}^{(t)}$ and $\{\widehat{\boldsymbol{C}}_k^{(t)}\}_{k=1}^K$, which are obtained in an alternating fashion using the offline "batch" approach, the general description of the proposed online adaptive QML solution algorithm is as follows:

Upon the reception of a new measurement $\boldsymbol{x}[t+1]$,
(i) Estimate the sources: $\widehat{\boldsymbol{S}}_t^{(t+1)} \triangleq \widehat{\boldsymbol{B}}^{(t)}\boldsymbol{X}^{(t+1)}$;
(ii) Using $\widehat{\boldsymbol{S}}_t^{(t+1)}$ and $\{\widehat{\boldsymbol{C}}_k^{(t)}\}$, estimate $\{\widehat{\boldsymbol{C}}_k^{(t+1)}\}$;
(iii) Using $\{\widehat{\boldsymbol{C}}_k^{(t+1)}\}$ and $\boldsymbol{X}^{(t+1)}$, construct $\{\widehat{\boldsymbol{Q}}_k^{(t+1)}\}$;
(iv) Using $\{\widehat{\boldsymbol{Q}}_k^{(t+1)}\}$, compute $\widehat{\boldsymbol{B}}^{(t+1)}$ as a solution of (2);
(v) Re-estimate the sources: $\widehat{\boldsymbol{S}}_{t+1}^{(t+1)} \triangleq \widehat{\boldsymbol{B}}^{(t+1)}\boldsymbol{X}^{(t+1)}$;

where $\widehat{\boldsymbol{Q}}_k^{(t)} \triangleq \frac{1}{t}\boldsymbol{X}^{(t)}\left(\widehat{\boldsymbol{C}}_k^{(t)}\right)^{-1}\boldsymbol{X}^{(t)\mathrm{T}}$ for every $k \in \{1,\ldots,K\}$. when the estimates of $\{\boldsymbol{C}_k\}_{k=1}^K$ are consistent[3] and all the initial estimates are "close" enough to their true values, the above procedure yields consistent estimates of the sources, by virtue of the continuous mapping theorem [18]. Note that $\widehat{\boldsymbol{S}}_t^{(t+1)}$ in step (i) can be readily constructed from $\widehat{\boldsymbol{S}}_{t+1}^{(t+1)}$ obtained in step (v) of the previous iteration, simply by concatenating the vector $\widehat{\boldsymbol{B}}^{(t)}\boldsymbol{x}[t+1]$ to its right. Relating to computational efficiency, if $t \gg 1$, the update in the target-matrices would be "small", and accordingly the update in the demixing-matrix' estimate would also be "small". Therefore, the computation of the new SeDJoCo solution may be obtained by Newton's method with Conjugate Gradient (NCG) efficiently (see [12] for further details), initialized by the current estimate, which effectively guarantees convergence in only a few single iterations. In addition, notice that since $\widehat{\boldsymbol{B}}^{(t+1)}$ is obtained by solving (2) with target-matrices $\{\widehat{\boldsymbol{Q}}_k^{(t+1)}\}$, computed using $\boldsymbol{X}^{(t+1)}$ and $\{\widehat{\boldsymbol{C}}_k^{(t+1)}\}_{k=1}^K$, it suffices to compute only the necessary update from the previously computed target-matrices $\{\widehat{\boldsymbol{Q}}_k^{(t)}\}$, and then, using the updated target-matrices, to compute the updated demixing-matrix estimate, as we demonstrate in detail for a simple example in the following subsection.

A. The Solution for First-Order Autoregressive Sources

Due to space limitations and for simplicity of the analytical derivation, we demonstrate here how the proposed solution scheme is applied to the case where the sources are all zero-mean first-order autoregressive (AR(1)) processes. Nevertheless, we note that the general and main principles of our proposed online adaptive solution are simply- and well-demonstrated by this test case.

Since the sources are all zero-mean AR(1), and assuming for simplicity that their generating driving-noise processes are all unit variance, their spectra, as well as their temporal covariance matrices $\{\boldsymbol{C}_k\}_{k=1}^K$, are each characterized by a single parameter, denoted here as $\{\alpha_k\}_{k=1}^K$, respectively. Thus, the (τ_1,τ_2)-th element of the temporal covariance matrix of the k-th source reads

$$C_k^{(t)}[\tau_1,\tau_2] = \frac{(\alpha_k)^{|\tau_1-\tau_2|}}{1-\alpha_k^2},\ \forall k \in \{1,\ldots,K\}, \tag{4}$$

for every $\tau_1,\tau_2 \in \{1,\ldots,t\}$. By simple computation, the elements of the inverse of the temporal covariance matrices, denoted for simplicity as $\boldsymbol{P}_k^{(t)} \triangleq \left(\boldsymbol{C}_k^{(t)}\right)^{-1}$ for every $k \in \{1,\ldots,K\}$, are given by

$$P_k^{(t)}[\tau_1,\tau_2] = \begin{cases} 1, & \tau_1=\tau_2=1,t \\ 1+\alpha_k^2, & \tau_1=\tau_2\neq 1,t \\ -\alpha_k, & |\tau_1-\tau_2|=1 \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

Now, expanding $Q_k^{(t)}[i,j]$ according to (3), we obtain

$$\begin{aligned} Q_k^{(t)}[i,j] &= \left(\frac{t-1}{t}\right)Q_k^{(t-1)}[i,j] + \\ &\quad \frac{1}{t}\boldsymbol{e}_i^{\mathrm{T}}\left[\boldsymbol{\mathcal{X}}^{(t)}[0] - \alpha_k\cdot\left(\boldsymbol{\mathcal{X}}^{(t)}[1]+\boldsymbol{\mathcal{X}}^{(t)}[1]^{\mathrm{T}}\right)\right]\boldsymbol{e}_j \quad (6) \\ &\triangleq \left(1-\frac{1}{t}\right)Q_k^{(t-1)}[i,j] + \left(\frac{1}{t}\right)\Delta Q_k^{(t)}[i,j], \quad (7) \end{aligned}$$

where we have defined $\{\boldsymbol{\mathcal{X}}^{(t)}[\ell] \triangleq \boldsymbol{x}[t]\boldsymbol{x}[t-\ell]^{\mathrm{T}}\}_{\ell=0}^1$ and

$$\Delta\boldsymbol{Q}_k^{(t)} \triangleq \boldsymbol{\mathcal{X}}^{(t)}[0] - \alpha_k\cdot\left(\boldsymbol{\mathcal{X}}^{(t)}[1]+\boldsymbol{\mathcal{X}}^{(t)}[1]^{\mathrm{T}}\right). \tag{8}$$

Hence, we may write in matrix form

$$\boldsymbol{Q}_k^{(t)} = \left(1-\frac{1}{t}\right)\boldsymbol{Q}_k^{(t-1)} + \left(\frac{1}{t}\right)\Delta\boldsymbol{Q}_k^{(t)}, \tag{9}$$

so the target-matrices may be simply updated by computation of the "update" matrices given in (8), followed by an update of the target-matrices, according to the resulting time-varying weighting which stems from the update equation (9). When α_k is replaced by its estimate $\widehat{\alpha}_k^{(t)}$ in (8), $\boldsymbol{Q}_k^{(t)}, \boldsymbol{Q}_k^{(t-1)}$ and $\Delta\boldsymbol{Q}_k^{(t)}$ are replaced by $\widehat{\boldsymbol{Q}}_k^{(t)}, \widehat{\boldsymbol{Q}}_k^{(t-1)}$ and $\Delta\widehat{\boldsymbol{Q}}_k^{(t)}$, respectively, in (9). Now, assuming $\widehat{\boldsymbol{B}}^{(t)}$ and $\widehat{\boldsymbol{\alpha}}^{(t)} \triangleq \left[\widehat{\alpha}_1^{(t)}\cdots\widehat{\alpha}_K^{(t)}\right]^{\mathrm{T}} \in \mathbb{R}^{K\times 1}$ are initial estimates for a sample size t, according to step (i), upon the arrival of $\boldsymbol{x}[t+1]$ we have $\widehat{\boldsymbol{S}}_t^{(t+1)}$. In step (ii), computation of $\widehat{\boldsymbol{\alpha}}^{(t+1)}$ based on $\widehat{\boldsymbol{\alpha}}^{(t)}$ and $\widehat{\boldsymbol{S}}_t^{(t+1)}$ is required. For the case of AR sources, consistent estimates of the AR parameters may be obtained by the Yule-Walker (Y-W) equations [19], [20] (which are also the asymptotic likelihood equations when the sources are Gaussian). In particular, for AR(1) sources the Y-W equations solutions are given (in matrix form) by

$$\widehat{\boldsymbol{\alpha}}^{(t)} = \mathrm{diag}\left(\widehat{\boldsymbol{R}}_s^{(t)}[1]\right)./\mathrm{diag}\left(\widehat{\boldsymbol{R}}_s^{(t)}[0]\right), \tag{10}$$

where the operator diag$(\cdot)$ returns the diagonal elements of its (square) matrix argument as a column vector, ./ denotes the element-wise division operation and $\widehat{\boldsymbol{R}}_s^{(t)}[\ell] \triangleq \widehat{\boldsymbol{S}}_t^{(t)}\boldsymbol{L}_\ell^{(t)}\left(\widehat{\boldsymbol{S}}_t^{(t)}\right)^{\mathrm{T}}$ for $\ell = 0,1$ where $\boldsymbol{L}_0^{(t)} \triangleq \frac{1}{t}\boldsymbol{I}$ and $\boldsymbol{L}_1^{(t)}$

[3] which may be any chosen SOS consistent estimation method

is the all-zeros matrix with $\frac{1}{2(t-1)}$ along its ± 1-th diagonals. Now, if we denote

$$\widehat{\boldsymbol{B}}^{(t)} = \widehat{\boldsymbol{B}}^{(t-1)} + \Delta\widehat{\boldsymbol{B}}^{(t)} \triangleq \left(\boldsymbol{I} + \boldsymbol{\mathcal{E}}^{(t)}\right)\widehat{\boldsymbol{B}}^{(t-1)}, \qquad (11)$$

where $\boldsymbol{\mathcal{E}}^{(t)} \triangleq \Delta\widehat{\boldsymbol{B}}^{(t)}\left(\widehat{\boldsymbol{B}}^{(t-1)}\right)^{-1}$, expanding $\widehat{\boldsymbol{R}}_s^{(t+1)}[\ell]$ yields (for $\ell = 0, 1$)

$$\widehat{\boldsymbol{R}}_s^{(t+1)}[\ell] = \frac{1}{t+1-\ell}\left(\boldsymbol{I} + \boldsymbol{\mathcal{E}}^{(t)}\right)\cdot \left((t-\ell)\cdot\widehat{\boldsymbol{R}}_s^{(t)}[\ell] + \widehat{\boldsymbol{\sigma}}_s^{(t|t+1)}[\ell]\right)\left(\boldsymbol{I} + \boldsymbol{\mathcal{E}}^{(t)}\right)^{\mathrm{T}}, \qquad (12)$$

where we have defined (for $\ell = 0, 1$)

$$\widehat{\boldsymbol{\sigma}}_s^{(\tau_1|\tau_2)}[\ell] \triangleq \frac{1}{2}\left(\widehat{\boldsymbol{s}}^{(\tau_1)}[\tau_2]\widehat{\boldsymbol{s}}^{(\tau_1)}[\tau_2-\ell]^{\mathrm{T}} + \widehat{\boldsymbol{s}}^{(\tau_1)}[\tau_2-\ell]\widehat{\boldsymbol{s}}^{(\tau_1)}[\tau_2]^{\mathrm{T}}\right). \qquad (13)$$

Thus, according to (10), the updated estimate $\widehat{\boldsymbol{\alpha}}^{(t+1)}$ is computed from $\{\widehat{\boldsymbol{R}}_s^{(t+1)}[\ell]\}_{\ell=0}^{1}$, which are simple updates of $\{\widehat{\boldsymbol{R}}_s^{(t)}[\ell]\}_{\ell=0}^{1}$ via (12), as a function of $\{\widehat{\boldsymbol{\sigma}}_s^{(t|t+1)}[\ell]\}_{\ell=0}^{1}$, respectively, and $\boldsymbol{\mathcal{E}}^{(t+1)}$. With that, we have completed deriving all the required update formulas for the implementation of the proposed online adaptive QML solution scheme in our scenario.

Summarizing the above, given $\widehat{\boldsymbol{B}}^{(t)}$ and $\widehat{\boldsymbol{\alpha}}^{(t)}$,

Upon the reception of a new measurement $\boldsymbol{x}[t+1]$,

(i) Estimate the sources: $\widehat{\boldsymbol{S}}_t^{(t+1)} \triangleq \widehat{\boldsymbol{B}}^{(t)}\boldsymbol{X}^{(t+1)}$;

(ii) Estimate $\{\widehat{\boldsymbol{\alpha}}_k^{(t+1)}\}$ by updating $\{\widehat{\boldsymbol{\alpha}}_k^{(t)}\}$ via (10):
  a) Construct $\{\widehat{\boldsymbol{\sigma}}_s^{(t|t+1)}[\ell]\}_{\ell=0}^{1}$;
  b) Update $\{\widehat{\boldsymbol{R}}_s^{(t)}[\ell]\}_{\ell=0}^{1}$ to $\{\widehat{\boldsymbol{R}}_s^{(t+1)}[\ell]\}_{\ell=0}^{1}$ via (12).

(iii) Using $\{\widehat{\boldsymbol{\alpha}}_k^{(t+1)}\}$ and $\{\boldsymbol{\mathcal{X}}^{(t+1)}[\ell]\}_{\ell=0}^{1}$, update $\{\widehat{\boldsymbol{Q}}_k^{(t)}\}$ to $\{\widehat{\boldsymbol{Q}}_k^{(t+1)}\}$ via (8) and (9);

(iv) Using $\{\widehat{\boldsymbol{Q}}_k^{(t+1)}\}$, compute $\widehat{\boldsymbol{B}}^{(t+1)}$ as a solution of (2) via NCG initialized by $\widehat{\boldsymbol{B}}^{(t)}$;

(v) Re-estimate the sources: $\widehat{\boldsymbol{S}}_{t+1}^{(t+1)} \triangleq \widehat{\boldsymbol{B}}^{(t+1)}\boldsymbol{X}^{(t+1)}$.

An initial estimate of $\boldsymbol{\alpha}$ may be set arbitrarily as some "educated" guess and an initial estimate of $\boldsymbol{B}$ for the online solution scheme is obtained as a "batch" SeDJoCo solution after an accumulation of a sufficiently large sample size, such that the subsequent updates will be small enough relative to its initial estimate's value.

IV. Simulation Results

Our proposed algorithm is demonstrated for two scenarios of a simulation experiment with the following setup. A set of $K = 3$ AR(1) sources was generated by filtering 3 statistically independent, zero-mean, unit-variance, temporally-i.i.d. noise-processes ("driving noise") in 3 different filters, determined by the AR(1) parameters presented in Table I. In the first scenario all driving noise signals were Gaussian, whereas in the second scenario they were all uniformly distributed. In each trial the mixing matrix' elements were drawn independently from a standard Gaussian distribution, the AR(1) parameters initial estimates were set arbitrarily to $\widehat{\boldsymbol{\alpha}}_0 = [0.8\ 0.6\ 0.4]^{\mathrm{T}}$ and the initial QMLE was obtained for a sample size of $T = 100$. Prior to the online operation, we ran two successive estimation of the sources and the AR(1) parameters (based only on the initial batch of measurements). Empirical results were obtained by 5000 independent trails.

AR(1) Parameters

α_1	α_2	α_3
0.91	-0.65	0.28

TABLE I: AR(1) parameters of the sources.

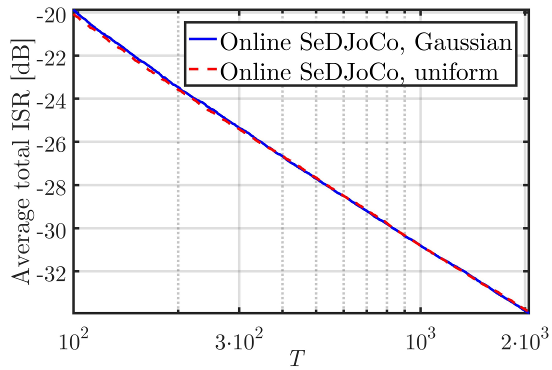

Fig. 1: Average total ISR vs. T with AR(1) sources for the two scenarios. Scenario 1: Gaussian driving-noises, Scenario 2: Uniform driving-noises. Results are based on 5000 independent trials

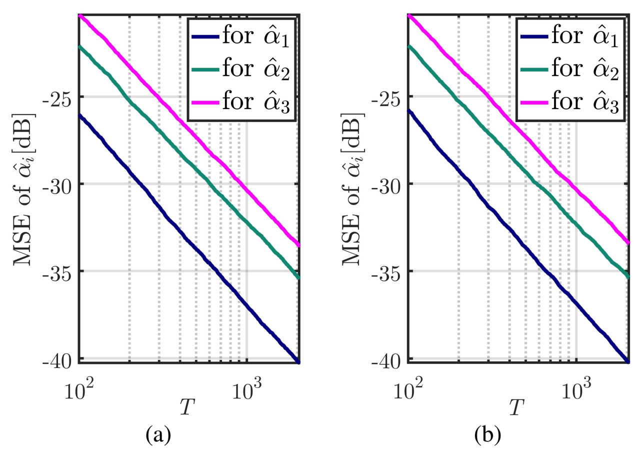

Fig. 2: Empirical MSE of the AR(1) parameters vs. T for the two scenarios. (a) Scenario 1: Gaussian driving-noises (b) Scenario 2: Uniform driving-noises. Results are based on 5000 independent trials.

Fig. 1 shows the average empirical total ISR versus T, obtained by the online SeDJoCo algorithm, for the two scenarios. Clearly, the ISR decreases monotonically and a consistency trend is evident. Furthermore, we observe that asymptotically the ISR does not depend on the sources' true distributions, in agreement with the theorem presented in [17]. Fig. 2 presents the empirical Mean Squared Error (MSE) in the estimation of the AR(1) parameters. It is easily seen that these estimates are

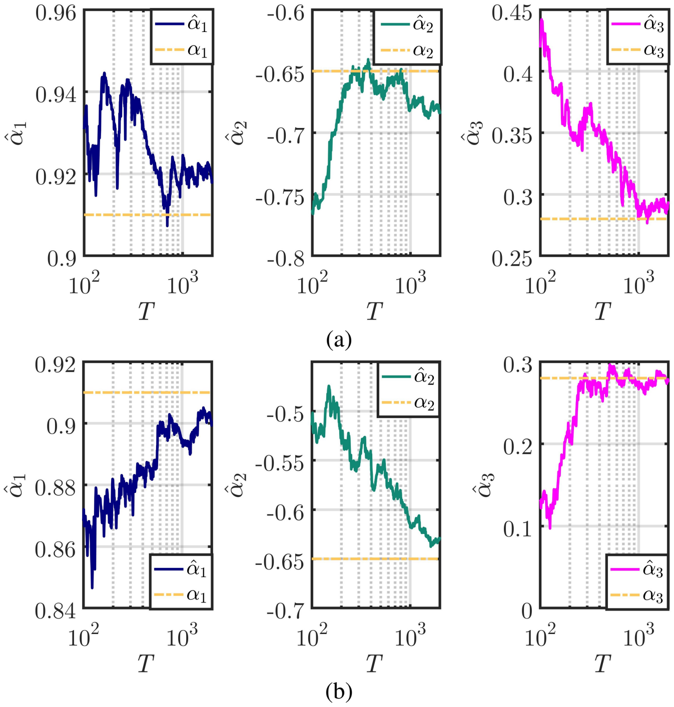

Fig. 3: The AR(1) parameters estimates (solid lines) vs. T and their true values (dashed lines) for the two scenarios in a (single) typical run of the algorithm. (a) Scenario 1: Gaussian driving-noises (b) Scenario 2: Uniform driving-noises. Note that differences between the two parts are mainly due to the differences in the sources' realizations, rather than in their statistics.

consistent, exhibiting (asymptotically) a constant rate of decay of their empirical MSEs. Fig. 3 presents the AR(1) parameters estimates values vs. T in a single run of the algorithm as a representative typical example of the convergence pattern.

V. Conclusion

We presented a SOS-based online adaptive QML algorithm for separation of stationary sources. Based on the Gaussian QML approach and a parametric model for the sources' spectra, the proposed algorithm alternates between updates of the spectra-characterizing parameters' estimates and the demixing-matrix' estimate at each iteration. The algorithm yields consistent separation due to the consistency of Gaussian QMLE. Moreover, it is computationally efficient due to the NCG computational efficiency and the update formulas we derived, which enable simple and fast updates of the estimates already obtained upon the arrival of a new measurement sample-vector.

We note in passing that such a scheme may also offer the possibility to track slow changes in the statistical properties of the sources and/or in the mixing matrix. Furthermore, in semi-blind scenarios, when the model precisely describes the true state of nature but the covariance values are still unknown, the proposed scheme would asymptotically reach optimal separation, i.e., attain the induced Cramér-Rao lower bound (iCRLB, [9]).

VI. Acknowledgment

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

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] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] A. Weiss, A. Yeredor, S. A. Cheema, and M. Haardt, "Performance Analysis of the Gaussian Quasi-Maximum Likelihood Approach for Independent Vector Analysis," submitted to IEEE Trans. on Signal Processing, 2018.

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

[19] G. Udny Yule, "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers," *Philosophical Transactions of the Royal Society of London Series A*, vol. 226, pp. 267–298, 1927.

[20] G. T. Walker, "On periodicity in series of related terms," *Proc. R. Soc. Lond. A*, vol. 131, no. 818, pp. 518–532, 1931.