Fast Approximate Joint Diagonalization Incorporating Weight Matrices

Petr Tichavský, *Senior Member, IEEE*, and Arie Yeredor, *Senior Member, IEEE*

***Abstract*—We propose a new low-complexity approximate joint diagonalization (AJD) algorithm, which incorporates nontrivial block-diagonal weight matrices into a weighted least-squares (WLS) AJD criterion. Often in blind source separation (BSS), when the sources are nearly separated, the optimal weight matrix for WLS-based AJD takes a (nearly) block-diagonal form. Based on this observation, we show how the new algorithm can be utilized in an iteratively reweighted separation scheme, thereby giving rise to fast implementation of asymptotically optimal BSS algorithms in various scenarios. In particular, we consider three specific (yet common) scenarios, involving stationary or block-stationary Gaussian sources, for which the optimal weight matrices can be readily estimated from the sample covariance matrices (which are also the target-matrices for the AJD). Comparative simulation results demonstrate the advantages in both speed and accuracy, as well as compliance with the theoretically predicted asymptotic optimality of the resulting BSS algorithms based on the weighted AJD, both on large scale problems with matrices of the size 100×100.**

***Index Terms*—Approximate joint diagonalization (AJD), autoregressive processes, blind source separation (BSS), nonstationary random processes.**

I. Introduction

THE problem of approximate joint diagonalization (AJD) of a set of matrices is frequently encountered in the context of blind source separation (BSS), and, more generally, in the field of multivariate statistical signal processing, whenever it is desired to fit a set of square, symmetric, real-valued,[1] $d \times d$ matrices $\widehat{\mathbf{R}}_{\mathbf{x}}[m]$, $m = 0, \ldots, M-1$ by structured matrices of the form

$$\mathbf{R}_{\mathbf{x}}[m] = \mathbf{A}_0 \mathbf{R}_{\mathbf{s}}[m] \mathbf{A}_0^T. \tag{1}$$

Manuscript received May 05, 2008; revised October 17, 2008. First published November 11, 2008; current version published February 13, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jonathon Chambers. This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572. A portion of this work was presented at the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, March 30–April 4, 2008.

P. Tichavský is with the Institute of Information Theory and Automation, 182 08 Prague 8, Czech Republic (e-mail: tichavsk@utia.cas.cz; jnielsen@utia.cas.cz).

A. Yeredor 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.2008.2009271

[1]Throughout this paper all matrices are assumed to be real-valued. Possible extension to the complex-valued case will be mentioned in our conclusions in Section VII.

Here $\mathbf{R}_{\mathbf{s}}[m] = \operatorname{diag}\left\{r_s^{(1)}[m], r_s^{(2)}[m], \ldots, r_s^{(d)}[m]\right\}$ denote unknown diagonal matrices, and $\mathbf{A}_0$ denotes an unknown matrix, often termed the *mixing* matrix, for easy reference, whereas its inverse $\mathbf{V}_0 = \mathbf{A}_0^{-1}$ is termed the *demixing* matrix. These terms are associated with the common interpretation of $\mathbf{A}_0$ and $\mathbf{V}_0$ in the context of BSS, where an observed multivariate $(d \times 1)$ process $\mathbf{x}[n]$ is modeled as an unknown linear mixture of some d unknown sources $\mathbf{s}[n]$, *viz.* $\mathbf{x}[n] = \mathbf{A}_0 \mathbf{s}[n]$.

Typically in BSS, the diagonal matrices $\mathbf{R}_{\mathbf{s}}[m]$ contain some statistical or structural properties of the sources, e.g., correlation matrices at different lags; different cumulant-slice matrices; covariance matrices within different time intervals; time-frequency distributions at different time-frequency points, and more. The "*target matrices*" $\widehat{\mathbf{R}}_{\mathbf{x}}[m]$ usually denote estimates of similar matrices pertaining to the observed mixtures. The diagonality of the matrices $\mathbf{R}_{\mathbf{s}}[m]$, which is often (but not always) attributed to the statistical independence of the sources, serves as the key to identifiability of the mixing matrix $\mathbf{A}_0$ from the matrices $\widehat{\mathbf{R}}_{\mathbf{x}}[m]$.

Some of the earlier AJD algorithms (e.g., by Cardoso and Souloumiac [4]) assume that $\mathbf{A}_0$ (and $\mathbf{V}_0$) are unitary. While this may be a reasonable assumption whenever some sphering (spatial whitening) preprocessing is applied, such a combined operation was shown (e.g., [5]) to limit the resulting performance. Nevertheless, this approach has become nearly common-practice in many BSS applications, perhaps due to its conceptual and computational simplicity. For example, in biomedical applications, the second-order blind identification (SOBI) algorithm [3] based on unitary AJD [4] has recently gained popularity [13], [21], [24].

In recent years quite a few algorithms which relax the unitarity assumption have been proposed. Among these, current state-of-the-art algorithms which are considerably computationally efficient relative to others (especially in large-scale problems with $d \gg 2$) seem to be: Pham's Log-Likelihood based AJD [17] (termed LLAJD in here), which is further constrained by the requirement that $\widehat{\mathbf{R}}_{\mathbf{x}}[m]$ must all be positive definite; FFDIAG by Ziehe *et al.* [33]; QAJD by Vollgraf and Obermayer [30]; FAJD by Li and Zhang [16] and QRJ2D by Afsari [1].

A common approach to AJD is to minimize some off-diagonality criterion applied to the transformed set $\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T$, thereby obtaining an estimate $\widehat{\mathbf{V}}$ of the demixing matrix as the minimizer of this criterion. Usually (depending on the off-diagonality criterion), such an approach requires to constrain $\widehat{\mathbf{V}}$, so as to evade trivial minimization by down-scaling towards $\widehat{\mathbf{V}} = \mathbf{0}$. In [4], such a constraint is naturally entailed in the unitarity assumption. When this assumption is relaxed, some alternative, more artificial constraints are to be considered. One such

possible constraint, considered, e.g., in [6], is for each row of the estimated demixing matrix $\widehat{\mathbf{V}}$ to have unit Euclidean norm. Another possible constraint, proposed in QAJD [30] (and also applied in here), is that $\widehat{\mathbf{V}}\widehat{\mathbf{R}}_{\mathbf{x}}[0]\widehat{\mathbf{V}}^T$ must have an all-ones main diagonal. This constraint usually corresponds (in the BSS context) to some scaling constraint on the estimated sources. In particular, when $\widehat{\mathbf{R}}_{\mathbf{x}}[0]$ is the observations' empirical covariance matrix, such a constraint is rather plausible, since the implied scaling in such a case translates into a standard unit-power constraint on the estimated sources. Of course, for other types of $\widehat{\mathbf{R}}_{\mathbf{x}}[0]$, such a constraint may be rather arbitrary, and, in general, the results might depend on the particular choice of $\widehat{\mathbf{R}}_{\mathbf{x}}[0]$. Naturally, this constraint is only applicable if $\widehat{\mathbf{R}}_{\mathbf{x}}[0]$ is positive definite. Note that $\widehat{\mathbf{R}}_{\mathbf{x}}[0]$ may or may not be included in the set of "target matrices" $\{\widehat{\mathbf{R}}_{\mathbf{x}}[m]\}$ to be jointly diagonalized.

QAJD is based on constrained minimization of the criterion

$$C_{LS}(\mathbf{V}) = \sum_{m=0}^{M-1} \left\| \mathrm{off}\left(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T\right)\right\|_F^2 \tag{2}$$

where the operator "off" nullifies the diagonal elements of a matrix and "$\|\cdot\|_F$" stands for the Frobenius norm.

In FAJD [16], degeneracy of $\widehat{\mathbf{V}}$ is evaded by adding a penalty term (proportional to $\log|\det\mathbf{V}|$) to (2). A different off-diagonality criterion, which also does not require explicit constraints (since it is scale-invariant in $\mathbf{V}$), is used in LLAJD [17]

$$C_{LL}(\mathbf{V}) = \sum_{m=0}^{M-1} \log \frac{\det \mathrm{ddiag}\left(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T\right)}{\det\left(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T\right)} \tag{3}$$

(the operator "ddiag" nullifies the off-diagonal elements of a square matrix, $\mathrm{ddiag}(\mathbf{M}) = \mathbf{M} - \mathrm{off}(\mathbf{M})$). This criterion is meaningful only for positive definite target-matrices $\{\widehat{\mathbf{R}}_{\mathbf{x}}[m]\}$.

Another suitable AJD criterion which is scale-invariant in $\mathbf{V}$ was proposed in [1] and [2]

$$C_{J2}(\mathbf{V}) = \sum_{m=0}^{M-1} \left\| \widehat{\mathbf{R}}_{\mathbf{x}}[m] - \mathbf{V}^{-1}\mathrm{ddiag}\left(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T\right)\mathbf{V}^{-T}\right\|_F^2. \tag{4}$$

In this paper, we propose an AJD approach which is based on a weighted least squares (WLS) criterion, to be presented in the sequel. However, our proposed algorithm does not minimize this WLS criterion directly. Using Gauss iterations (e.g., [22]) in a specially adapted, computationally efficient form, we apply successive diagonalizing ("demixing") transformations to the target matrices. The process proceeds until the transformed target matrices reach a form for which the direct ("mixing") minimizer of the WLS criterion is the identity matrix (we elaborate on "direct" versus "indirect" minimization in Section II).

In addition to its computational efficiency, our approach offers the possibility to incorporate proper weighting in the WLS criterion, which is useful in many BSS scenarios (see, e.g., [8], [23], [25], and [31]). Such weighting can improve (or even optimize, asymptotically) the performance of AJD-based BSS approaches by accounting for statistical properties of the estimated target set.

Our algorithm is given the acronym WEDGE (Weighted Exhaustive Diagonalization with Gauss itErations). As we shall show, in its unweighted (or uniformly weighted) version (termed U-WEDGE),[2] our solution is closely related to both FFDIAG and QAJD.

In order to enable comparison of the weighted version, we also considered possible modification of QAJD to allow weighting.[3] While the resulting Weighted QAJD (W-QAJD) is significantly more computationally intensive than WEDGE, we show that the small-errors perturbations of W-QAJD and WEDGE are similar.

The paper is organized as follows: In Section II we present our approach in its unweighted version first, for simplicity of the exposition. We then consider the weighted version in Section III and present our iterative algorithm in Section IV. Intricate weight matrices suitable for WEDGE are derived in Section V for three different BSS scenarios. Computer simulations in Section VI demonstrate the validity of our analysis and exhibit the computational properties and accuracy performance of WEDGE.

Our weighted version of QAJD is proposed in Appendix A, where we also deduce the small-errors similarity of W-QAJD to WEDGE.

II. Diagonalization With Uniform Weights

We consider the unweighted (or uniformly weighted) U-WEDGE first. As already implied in the Introduction, there are (at least) two possible ways to express the desired joint-diagonality property. One is a so-called "direct" form (used, e.g., in [29] and [32])

$$\widehat{\mathbf{R}}_{\mathbf{x}}[m] \approx \widehat{\mathbf{A}}\widehat{\mathbf{R}}_{\mathbf{s}}\widehat{\mathbf{A}}^T \tag{5}$$

and the other is the "indirect" form (used, e.g., in [4] and [30])

$$\widehat{\mathbf{V}}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\widehat{\mathbf{V}}^T \approx \widehat{\mathbf{R}}_{\mathbf{s}}[m] \tag{6}$$

(both for $m = 0, 1, \ldots, M-1$). When the joint diagonality is exact, (5) and (6) are obviously equivalent, with $\widehat{\mathbf{V}} = \widehat{\mathbf{A}}^{-1}$. However, when the relation is approximate, and some measure of the matrix fit is used for both (5) and (6), the matrix $\widehat{\mathbf{V}}$ which optimizes the fit in (6) will not necessarily be the inverse of the matrix $\widehat{\mathbf{A}}$ which optimizes the fit in (5) (except in some particular cases).

Suppose that we fuse these two forms into one, using a least-squares criterion for the matrix fit; Namely, for any two matrices $\mathbf{V}$ and $\mathbf{A}$ define

$$\widetilde{C}_{LS}(\mathbf{V},\mathbf{A}) \triangleq \sum_{m=0}^{M-1} \left\| \mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T - \mathbf{A}\mathbf{D}_{m,\mathbf{V}}\mathbf{A}^T\right\|_F^2 \tag{7}$$

where $\mathbf{D}_{m,\mathbf{V}} \triangleq \mathrm{ddiag}(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T)$. Now, for any "demixing" matrix $\mathbf{V}$, one can find a "mixing" matrix $\bar{\mathbf{A}}$ which minimizes $\widetilde{C}_{LS}(\mathbf{V},\mathbf{A})$ with respect to $\mathbf{A}$. This $\mathbf{A} = \bar{\mathbf{A}}$

[2] In [28] the algorithms WEDGE and U-WEDGE were called WAJD and UWAJD, respectively.

[3] Similar modification of FFDIAG is meaningless, since FFDIAG does not explicitly minimize any criterion in which the weighting can be incorporated.

can be considered the "residual mixing" matrix which remains after the "demixing" matrix $\mathbf{V}$ is applied to the target matrices. In other words, define

$$\begin{aligned}\boldsymbol{\Theta}(\mathbf{V}) &\triangleq \operatorname{argmin}_{\mathbf{A}} \widetilde{C}_{LS}(\mathbf{V},\mathbf{A}) \\ &= \operatorname{argmin}_{\mathbf{A}} \sum_{m=0}^{M-1} \left\| \mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T - \mathbf{A}\mathbf{D}_{m,\mathbf{V}}\mathbf{A}^T \right\|_F^2 .\end{aligned} \tag{8}$$

The (matrix) function $\boldsymbol{\Theta}(\mathbf{V})$ fits to any "demixing" matrix $\mathbf{V}$ its "residual mixing" matrix $\bar{\mathbf{A}}$, which attains the best "direct" LS fit of the transformed target matrices.

Suppose now, that the matrix $\mathbf{V}$ solves the equation $\boldsymbol{\Theta}(\mathbf{V}) = \mathbf{I}$. Roughly speaking, this implies that the set of matrices $\left\{\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T\right\}$ cannot be jointly diagonalized any more, since its "residual mixing" matrix, or its "best direct-form diagonalizer" (in the LS sense) is $\bar{\mathbf{A}} = \mathbf{I}$, the identity matrix.[4]

We show in Appendix B, that a necessary condition for $\boldsymbol{\Theta}(\mathbf{V}) = \mathbf{I}$ is a simpler set of nonlinear "normal equations"

$$\begin{aligned}\boldsymbol{\Psi}(\mathbf{V}) &\triangleq \operatorname{off}\left[\sum_{m=0}^{M-1} \mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T \cdot \operatorname{ddiag}\left(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T\right)\right] \\ &= \mathbf{0}.\end{aligned} \tag{9}$$

Note that some rows-scaling constraint for $\mathbf{V}$ is still required, since if any matrix $\mathbf{V}$ solves the set (9), so does the matrix $\mathbf{DV}$ where $\mathbf{D}$ is any diagonal matrix. This can be easily observed by noting that for any diagonal matrix $\mathbf{D}$, we have $\boldsymbol{\Psi}(\mathbf{DV}) = \mathbf{D}\boldsymbol{\Psi}(\mathbf{V})\mathbf{D}^3$. This rows-scaling invariance in $\mathbf{V}$ is also evident from the basic (8), as well as from the well-known inherent scale-ambiguity in BSS. We, therefore, also employ a scaling convention, which is the one used in QAJD—$\operatorname{ddiag}(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[0]\mathbf{V}^T) = \mathbf{I}$. For convenience in the notations we shall assume that the matrix $\widehat{\mathbf{R}}_{\mathbf{x}}[0]$ is part of the target-set, hence, all summations indices over m begin at $m = 0$. When $\widehat{\mathbf{R}}_{\mathbf{x}}[0]$ is to be excluded from the set, the summation should begin at $m = 1$.

It is interesting to note here, that the AJD solution provided by the FFDIAG algorithm [33] can also be shown to satisfy our condition (9).[5] Up to date it has been unclear whether or not FFDIAG actually minimizes any explicit criterion or solves any explicit equation (expressed in terms of the target-matrices). Our observation identifies such an equation, and, moreover, asserts that any solution of (9) is also a stationary point of FFDIAG, and vice versa. In this sense, results produced by FFDIAG and U-WEDGE are equivalent.

Note further, that (9) is only a necessary condition for a solution of (8), since it only implies that $\mathbf{A} = \mathbf{I}$ is a stationary point, but not necessarily a minimum, of the LS criterion in (8). Thus, while any solution of FFDIAG solves (9) as well, it is not guaranteed to also solve (8), namely to be a U-WEDGE solution. Observe, for example, that given any set of target matrices, one can usually construct two additional symmetric matrices, such that when these matrices are appended to the set, $\mathbf{V} = \mathbf{I}$ solves (9), see Appendix C for more details. However, such an undesired solution is usually an unstable stationary point: with any slight perturbation thereof, U-WEDGE would rapidly drift to another solution of (8), providing a reasonable diagonalization solution.

Also note that the other algorithms may also have similar undesired solutions. For example, the constrained criterion (2) of QAJD appears to have local minima at matrices that are close to false solutions of (9), see the discussion in [16]. Similarly, as shown in [2], the LLAJD criterion (3) may exhibit false local minima or saddle points at matrices satisfying

$$\frac{1}{M}\sum_{m=0}^{M-1} \mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T \left(\operatorname{ddiag}(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T)\right)^{-1} = \mathbf{I}. \tag{10}$$

In Appendix A we shall observe another interesting relation to another AJD algorithm, namely QAJD. Unlike FFDIAG, QAJD minimizes an explicit LS criterion (2), which can be modified into a WLS criterion. To enable a more general comparison between WEDGE and a weighted version of QAJD, we shall develop (in Appendix A) a weighted version of QAJD, W-QAJD (which is considerably more computationally intense than WEDGE). As a byproduct, we shall be able to observe similarity in the small-errors perturbations between W-QAJD and WEDGE (which would obviously also hold for the particular cases of unweighted QAJD and U-WEDGE).

Indeed, numerical simulations show, that the constrained minimization of $C_{LS}(\mathbf{V})$ in (2) (i.e., outcome of QAJD) and the similarly constrained (scaled) solution to $\boldsymbol{\Theta}(\mathbf{V}) = \mathbf{I}$ (outcome of WEDGE) are quite often very close to each other, especially when the target matrices consist of small perturbations of a set which is exactly jointly diagonalizable. However, this is certainly not always the case in general, as evident in the simulation results which will be presented in the sequel.

III. Incorporating a Weight Matrix

The unweighted QAJD and U-WEDGE can be generalized by introducing an arbitrary positive definite weight matrix $\mathbf{W}$. To this end, in the generalized QAJD, the LS criterion $C_{LS}(\mathbf{V})$ in (2) would be replaced by a quadratic form of off-diagonal elements of $\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T$

$$C_{WLS}(\mathbf{V}) = [\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V})]^T \mathbf{W} \widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V}) \tag{11}$$

where $\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V})$ is an $Md(d-1)/2$-dimensional column vector composed of all off-diagonal elements of $\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T$ below the main diagonal of the matrices for $m = 0, \ldots, M-1$. For convenience in some of the subsequent derivations, we prefer to group the elements of $\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V})$ according to their locations in the matrices (rather than according to their matrix-affiliation), namely

$$\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V}) \triangleq \left[\widehat{\mathbf{r}}_{21}^T, \widehat{\mathbf{r}}_{31}^T, \ldots, \widehat{\mathbf{r}}_{d1}^T, \widehat{\mathbf{r}}_{32}^T, \ldots, \widehat{\mathbf{r}}_{d2}^T, \ldots, \widehat{\mathbf{r}}_{d,d-1}^T\right]^T \tag{12}$$

[4]Note that the minimized expression in (8) is insensitive to the signs of the columns of $\bar{\mathbf{A}}$—we, therefore, employ a convention by which argmin always selects the minimizer with nonnegative diagonal elements.

[5]To observe this, note that if (and only if) (9) is satisfied, all of the $y_{i,j}$ terms on the bottom of [33, p. 783] (an unnumbered equation in there) vanish, and the update process in FFDIAG is thereby halted.

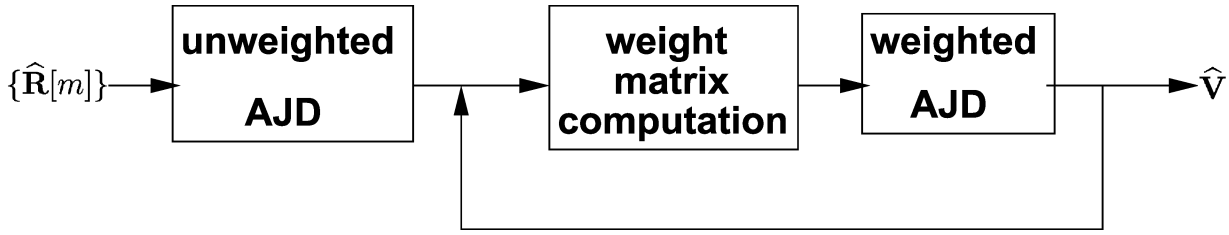

Fig. 1. A general separation scheme with iteratively estimated weights.

where

$$\widehat{\mathbf{r}}_{k\ell}(\mathbf{V}) \triangleq \left[\left(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[0]\mathbf{V}^T\right)_{k\ell},\ldots,\left(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[M-1]\mathbf{V}^T\right)_{k\ell}\right]^T \tag{13}$$

are $M\times 1$ vectors [the arguments $(\mathbf{V})$ are omitted for brevity in (12)].

Similar weighting can be applied for turning U-WEDGE into WEDGE, by replacing $\widetilde{C}_{LS}$ of (7) with a weighted version using a weight-matrix $\mathbf{W}$

$$\widetilde{C}_{WLS}(\mathbf{V},\mathbf{A}) \triangleq [\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V})-\mathbf{f}(\mathbf{A})]^T\mathbf{W}[\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V})-\mathbf{f}(\mathbf{A})] \tag{14}$$

where $\mathbf{f}(\mathbf{A})$ is a suitable quadratic form of the off-diagonal elements of $\mathbf{A}$. More specifically

$$\mathbf{f}(\mathbf{A}) \triangleq \left[\mathbf{f}_{21}^T,\mathbf{f}_{31}^T,\ldots,\mathbf{f}_{d1}^T,\mathbf{f}_{32}^T,\ldots,\mathbf{f}_{d2}^T,\ldots,\mathbf{f}_{d,d-1}^T\right]^T \tag{15}$$

where

$$\begin{aligned}\mathbf{f}_{k\ell}(\mathbf{A}) &\triangleq \left[\left(\mathbf{A}\mathbf{D}_{0,\mathbf{V}}\mathbf{A}^T\right)_{k\ell},\ldots,\left(\mathbf{A}\mathbf{D}_{M-1,\mathbf{V}}\mathbf{A}^T\right)_{k\ell}\right]^T\\ \mathbf{D}_{m,\mathbf{V}} &\triangleq \mathrm{ddiag}\left(\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T\right)\end{aligned} \tag{16}$$

for $k,\ell=1,\ldots,d$, $k>\ell$, and $m=0,1,\ldots,M-1$.

The equation $\boldsymbol{\Theta}(\mathbf{V})=\mathbf{I}$ is thereby simply replaced with $\boldsymbol{\Theta}_{\mathbf{W}}(\mathbf{V})=\mathbf{I}$, where

$$\begin{aligned}\boldsymbol{\Theta}_{\mathbf{W}}(\mathbf{V}) &\triangleq \mathrm{argmin}_{\mathbf{A}}\widetilde{C}_{WLS}(\mathbf{V},\mathbf{A})\\ &= \mathrm{argmin}_{\mathbf{A}}[\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V})-\mathbf{f}(\mathbf{A})]^T\mathbf{W}\\ &\quad\times[\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V})-\mathbf{f}(\mathbf{A})].\end{aligned} \tag{17}$$

Note that $\widetilde{C}_{WLS}(\mathbf{V},\mathbf{A})$ above is not the most general weighted form of $\widetilde{C}_{LS}(\mathbf{V},\mathbf{A})$, since, due to the special structure of $\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V})$ and $\mathbf{f}(\mathbf{A})$, it only accounts for off-diagonal elements of the difference matrices in the quadratic form. This dissimilarity merely entails deliberate, simplifying elimination of redundant terms, which are usually meaningless in the context of approximate joint diagonalization, and are given null-weight in an optimally weighted blind separation scenario.

Weighted AJD is desirable whenever it is possible to characterize random variations of $\widehat{\mathbf{R}}_{\mathbf{s}}[m]\approx\mathbf{V}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\mathbf{V}^T$ around their theoretical counterparts $\mathbf{R}_{\mathbf{s}}[m]$ in terms of their first and second moments. In such cases, an optimal weight matrix $\mathbf{W}$ (in the sense of minimum mean square error in the estimation of $\mathbf{V}$) is defined (assuming small errors) as the inverse of the covariance matrix of $\widehat{\mathbf{r}}_{\mathbf{s}}$. However, a common problem in such cases is that since the statistics of the source signals are unknown, estimating the covariance of $\widehat{\mathbf{r}}_{\mathbf{s}}$ directly from the observed mixtures can be prohibitively complicated. However, in a near-separation condition, when each source is (nearly) individually available, the relevant statistical properties of each source can be estimated from the data, possibly leading to reliable estimates of the covariance of $\widehat{\mathbf{r}}_{\mathbf{s}}$. Then, the implied weight matrix can be used to attain improved separation, which would in turn yield improved estimates of $\widehat{\mathbf{r}}_{\mathbf{s}}$. The process may be iterated a few times. Such an approach was taken, for example, in the optimal fourth-order identification algorithm (OFORIA, [23]), where the target matrices were the covariance matrix and cumulant slices; and more recently in the fast implementation of "weights-adjusted SOBI" (WASOBI) [25]), where the target matrices were correlations at different time-lags. The estimation scheme is depicted in Fig. 1.

In this paper we shall not address the WLS criterion (11) in its full generality, but only in a form by which the weight matrix $\mathbf{W}$ is block diagonal, with $M\times M$ blocks $\mathbf{W}_{k\ell}$, $k,\ell=1,\ldots,d$, $k>\ell$, each corresponding to the respective $\widehat{\mathbf{r}}_{k\ell}$ (it is important to note that the notation $\mathbf{W}_{k\ell}$ does *not* refer to the (k,ℓ)th block of $\mathbf{W}$, but rather to a block along its diagonal, corresponding to the covariance of $\widehat{\mathbf{r}}_{k\ell}$). Due to the special structure (12), (13) of $\widehat{\mathbf{r}}_{\mathbf{s}}$, such block-diagonal weighting can be optimal whenever the vector-pairs $\widehat{\mathbf{r}}_{k\ell}$, $\widehat{\mathbf{r}}_{k'\ell'}$ are uncorrelated for all $(k,\ell)\neq(k',\ell')$ (although each vector $\widehat{\mathbf{r}}_{k\ell}$ may have, and usually has, correlated elements). Fortunately, it so happens in a BSS context, that due to the independence of the sources, such a block-decorrelation condition can often be encountered when the sources are nearly separated, as we shall show in the sequel.

The criterion (11) of the W-QAJD can then be expressed as

$$C_{WLS}(\widehat{\mathbf{V}}) = \sum_{k>\ell}^{d}[\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}})]^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}}). \tag{18}$$

Similarly, for WEDGE, $\boldsymbol{\Theta}_{\mathbf{W}}(\mathbf{V})$ of (17) can be expressed as

$$\begin{aligned}\boldsymbol{\Theta}_{\mathbf{W}}(\mathbf{V}) = \mathrm{argmin}_{\mathbf{A}}\sum_{k>\ell}^{d}&[\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}})-\mathbf{f}_{k\ell}(\mathbf{A})]^T\mathbf{W}_{k\ell}\\ &\cdot[\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}})-\mathbf{f}_{k\ell}(\mathbf{A})].\end{aligned} \tag{19}$$

By requiring that $\mathbf{A}=\mathbf{I}$ be a stationary point of the sum in (19), it can be shown that a necessary condition for $\widehat{\mathbf{V}}$ to be a solution of $\boldsymbol{\Theta}_{\mathbf{W}}(\mathbf{V})=\mathbf{I}$, is to also be a solution of the system of equations

$$\begin{aligned}[\widehat{\mathbf{r}}_{\ell\ell}(\mathbf{V})]^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{k\ell}(\mathbf{V}) &= 0\\ [\widehat{\mathbf{r}}_{kk}(\mathbf{V})]^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{k\ell}(\mathbf{V}) &= 0\end{aligned} \tag{20}$$

$k,\ell=1,\ldots,d$, $k>\ell$ [compare to (24) in Section IV].

IV. Fixed-Point Iteration for WEDGE

We shall now propose an iterative algorithm for solving $\boldsymbol{\Theta}_{\mathbf{W}}(\mathbf{V}) = \mathbf{I}$. As an initial estimate of $\mathbf{V}$ we begin with $\widehat{\mathbf{V}}^{[0]} = (\widehat{\mathbf{R}}_{\mathbf{x}}[0])^{-1/2}$, so as to satisfy the scaling constraint. Then, in each iteration i, we seek an estimate of the "residual mixing" matrix, $\widehat{\mathbf{A}}$ (the minimizer of the WLS criterion in (19)), so as to fit the partially diagonalized matrices $\widehat{\mathbf{R}}_{\mathbf{s}}[m] \triangleq \widehat{\mathbf{V}}^{[i-1]}\widehat{\mathbf{R}}_{\mathbf{x}}[m](\widehat{\mathbf{V}}^{[i-1]})^T$ by matrix products of the form $\widehat{\mathbf{A}}\mathbf{D}_{\mathbf{s}}[m]\widehat{\mathbf{A}}^T$, where $\mathbf{D}_{\mathbf{s}}[m] = \operatorname{ddiag}(\widehat{\mathbf{R}}_{\mathbf{s}}[m])$. Once $\widehat{\mathbf{A}}$ is found, the estimated demixing matrix is updated as $\widehat{\mathbf{V}}^{[i]} = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{V}}^{[i-1]}$. Next, although it is not strictly necessary, we recommend subsequently to normalize $\widehat{\mathbf{V}}^{[i]}$ by suitable scaling of its rows, like in QAJD, to fulfil the constraint

$$\operatorname{ddiag}\left(\widehat{\mathbf{V}}^{[i]}\widehat{\mathbf{R}}_{\mathbf{x}}[0](\widehat{\mathbf{V}}^{[i]})^T\right) = \mathbf{I}.$$

Any other reasonable way of normalization may work equally well.

In order to find $\widehat{\mathbf{A}}$ in each iteration, we apply Gauss' iterative method (e.g., [22]), a generic tool for minimization of a quadratic form which depends on a nonlinear function of the parameters. More specifically, we apply

$$\boldsymbol{\theta}^{[j+1]} = \boldsymbol{\theta}^{[j]} + \left[\mathbf{F}_j^T\mathbf{W}\mathbf{F}_j\right]^{-1}\mathbf{F}_j^T\mathbf{W}\left[\widehat{\mathbf{r}}_{\mathbf{s}}(\mathbf{V}) - \mathbf{f}(\boldsymbol{\theta}^{[j]})\right] \quad (21)$$

where j is the internal (nested) Gauss-iteration index, $\boldsymbol{\theta}^{[j]} = \operatorname{vec}(\mathbf{A}^{[j]})$, and $\mathbf{F}_j = \partial\mathbf{f}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[j]}}$ (assumed to have full rank). Here $\mathbf{f}(\boldsymbol{\theta})$ is the same as $\mathbf{f}(\mathbf{A})$ of (15) (although its argument has changed from the matrix $\mathbf{A}$ to its vectorized form $\boldsymbol{\theta}$). Exploiting the block-diagonality of $\mathbf{W}$, (21) can be rewritten as

$$\boldsymbol{\theta}^{[j+1]} = \boldsymbol{\theta}^{[j]} + \left\{\sum_{k>\ell}^{d}\mathbf{F}_{k\ell}^{[j]T}\mathbf{W}_{k\ell}\mathbf{F}_{k\ell}^{[j]}\right\}^{-1} \cdot\left\{\sum_{k>\ell}^{d}\mathbf{F}_{k\ell}^{[j]T}\mathbf{W}_{k\ell}\left(\widehat{\mathbf{r}}_{k\ell}(\mathbf{V}) - \mathbf{f}_{k\ell}(\boldsymbol{\theta}^{[j]})\right)\right\} \quad (22)$$

where $\mathbf{F}_{k\ell}^{[j]}$ is the derivative (matrix) of $\mathbf{f}_{k\ell}$ with respect to $\boldsymbol{\theta}$, whose elements can be easily shown to be given by

$$\left.\frac{\partial\mathbf{f}_{k\ell}(\boldsymbol{\theta})}{\partial\widehat{A}_{pq}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[j]}} = \left(\delta_{kp}\widehat{A}_{\ell q}^{[j]} + \delta_{\ell p}\widehat{A}_{kq}^{[j]}\right)\left(\mathbf{I}_M\otimes\mathbf{e}_q^T\right)\mathbf{d} \quad (23)$$

δ_{kp} denoting Kronecker's delta, $\mathbf{e}_q$ the qth column of $\mathbf{I}_d$, and $\mathbf{d}$ is a vector composed of diagonal elements of $\widehat{\mathbf{R}}_{\mathbf{s}}[m]$, $m = 0,\ldots,M-1$, namely

$$\mathbf{d} = \left[\operatorname{diag}(\widehat{\mathbf{R}}_{\mathbf{s}}[0])^T,\ldots,\operatorname{diag}(\widehat{\mathbf{R}}_{\mathbf{s}}[M-1])^T\right]^T.$$

Assume now, that the initial condition (in the internal Gauss iterations) for $\boldsymbol{\theta}$ is selected as $\widehat{\mathbf{A}}^{[0]} = \mathbf{I}$. The linear system for $\boldsymbol{\theta}^{[1]}$ is nicely decoupled in this case, so that elements of $\widehat{\mathbf{A}}^{[1]}$ can be obtained merely by solving the following $d(d-1)/2$ systems of 2×2 for $k,\ell = 1,\ldots,d,\ k>\ell$:

$$\begin{bmatrix}\widehat{A}_{k\ell}^{[1]}\\ \widehat{A}_{\ell k}^{[1]}\end{bmatrix} = \begin{bmatrix}\widehat{\mathbf{r}}_{\ell\ell}^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{\ell\ell} & \widehat{\mathbf{r}}_{kk}^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{\ell\ell}\\ \widehat{\mathbf{r}}_{kk}^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{\ell\ell} & \widehat{\mathbf{r}}_{kk}^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{kk}\end{bmatrix}^{-1}\begin{bmatrix}\widehat{\mathbf{r}}_{\ell\ell}^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{k\ell}\\ \widehat{\mathbf{r}}_{kk}^T\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{k\ell}\end{bmatrix} \quad (24)$$

where we have used the relation $\left(\mathbf{I}_M\otimes\mathbf{e}_q^T\right)\mathbf{d} = \widehat{\mathbf{r}}_{qq}$. Computed for all pairs (k,ℓ), (24) realizes one update of $\widehat{\mathbf{A}}$ at complexity $O(M^2d^2)$. In the case of uniform weights, the complexity is only $O(Md^2)$. As for the diagonal elements $\widehat{A}_{kk}^{[1]}$, the respective columns of the derivative matrices $\mathbf{F}_{k\ell}^{[0]}$ (at $\widehat{\mathbf{A}}^{[0]} = \mathbf{I}$) are all-zeros, therefore these elements remain unchanged from their initial values of 1.

Note that the equations in (20) are indeed obtained when requiring zero updates for all elements of $\widehat{\mathbf{A}}$ in (24), namely when the initial guess $\widehat{\mathbf{A}}^{[0]} = \mathbf{I}$ is the minimizer of the quadratic form, so $\boldsymbol{\Theta}_{\mathbf{W}}(\mathbf{V}) = \mathbf{I}$.

To proceed, rather than reemploy (22) directly for $j = 1,2,\ldots$ (which would no longer enjoy the decoupling induced by $\widehat{\mathbf{A}}^{[0]} = \mathbf{I}$), we simply use $\widehat{\mathbf{A}}^{[1]}$ to update $\widehat{\mathbf{V}}^{[i]}$ as $\widehat{\mathbf{V}}^{[i]} = (\widehat{\mathbf{A}}^{[1]})^{-1}\widehat{\mathbf{V}}^{[i-1]}$. In other words, we merely employ a single Gauss iteration nested within each outer iteration. Note that the computational efficiency of our approach dwells on our ability to use the identity matrix $\mathbf{I}$ as an initial guess for the Gauss step in each iteration—which decouples the solution of a high-dimensional system into the solution of several small-dimensional systems (24).

Convergence of the algorithm is nearly quadratic, as inherited from the Gauss iterations. Simulations confirm very good global convergence even in high dimensions, significantly outperforming the competitors in terms of speed (in Matlab), with similar estimation accuracy (in the unweighted version).

Note further, that the proposed algorithm can be easily modified for a tracking environment. Since one update of the demixing matrix is computationally very cheap, such updates can be interlaced with updates of the target matrices $\{\widehat{\mathbf{R}}_{\mathbf{x}}[m]\}$.

To conclude this section, we provide a "pseudocode" for WEDGE.

`Inputs`:
- M "`target matrices`": $\widehat{\mathbf{R}}_{\mathbf{x}}[m] \in \mathbb{R}^{d\times d}$, $m = 0,1,\ldots,M-1$;
- $d(d-1)/2$ "`weight matrices:`" $\mathbf{W}_{k\ell}\in\mathbb{R}^{M\times M}$, $k,\ell = 1,2,\ldots,d,\ k>\ell$
- $\widehat{\mathbf{V}}^{[0]}$—`an initial guess for` $\widehat{\mathbf{V}}$(`may be set to` $\widehat{\mathbf{V}}^{[0]} = (\widehat{\mathbf{R}}_{\mathbf{x}}[0])^{-1/2}$).

`Outputs`: $\widehat{\mathbf{V}}$—`the estimated unmixing matrix.`

`Proceed for` $i = 1,2,\ldots$ `until convergence:`

1) `Generate the transformed set` $\widehat{\mathbf{R}}_{\mathbf{s}}[m] = \widehat{\mathbf{V}}^{[i-1]}\widehat{\mathbf{R}}_{\mathbf{x}}[m](\widehat{\mathbf{V}}^{[i-1]})^T$(`for` $m = 0,1,\ldots M-1$);
2) `Form the` $d(d+1)/2$ `vectors` $\widehat{\mathbf{r}}_{k\ell} = \left[(\widehat{\mathbf{R}}_{\mathbf{s}}[0])_{k\ell}\ (\widehat{\mathbf{R}}_{\mathbf{s}}[1])_{k\ell}\ldots(\widehat{\mathbf{R}}_{\mathbf{s}}[M-1])_{k\ell}\right]^T$ (`for` $k,\ell = 0,1,\ldots,d,\ k\geq\ell$);
3) `Set` $\widehat{\mathbf{A}}^{[1]} = \mathbf{I}$ `and substitute all of its off-diagonal elements with solutions of the` $d(d-1)/2$ `systems of` 2×2 `equations in` (24) (`for` $k,\ell = 0,1,\ldots,d,\ k>\ell$);
4) `Set` $\widehat{\mathbf{V}}^{[i]} = (\widehat{\mathbf{A}}^{[1]})^{-1}\widehat{\mathbf{V}}^{[i-1]}$
5) `Re-normalize rows of` $\widehat{\mathbf{V}}^{[i]}$.

`Upon convergence` $(i = i_{\text{end}})$, `set` $\widehat{\mathbf{V}} = \widehat{\mathbf{V}}^{[i_{\text{end}}]}$.

V. Weight Matrices in Different BSS Scenarios

In this section, we provide three different examples for three different BSS scenarios, in which introduction of proper

weighting into the AJD process can attain asymptotically optimal separation results. The weighting is introduced into the AJD via the use of WEDGE in the iteratively reweighted scheme outlined in Fig. 1 above. All three scenarios assume the noiseless, static, invertible mixing model $\mathbf{x}[n] = \mathbf{A}_0\mathbf{s}[n]$ (for $n = 1, 2, \ldots, N$). The separation is based on second-order statistics only, and the sources are all assumed to be Gaussian—therefore, closed-form expressions for the (estimated) covariance of correlations estimates can be easily obtained and exploited for (asymptotically) optimal weighting.

In Section V-A, we consider the separation of stationary autoregressive (AR) sources. In Section V-B we consider separation of nonstationary sources, assumed to be block-stationary and white within blocks. In fact, under asymptotic conditions the resulting optimally weighted algorithm can be regarded in this case as a novel implementation of the block Gaussian likelihood algorithm (BGL, by Pham, [19]), since they are both asymptotically optimal. However, we also show (in Section V-C) how weighting can be exploited in the case of block-stationary sources which are not white within blocks—to which the BGL algorithm cannot be readily applied.

A. Stationary AR Sources: WASOBI

In this subsection, we derive weight matrices for WASOBI, for separation of independent stationary sources, modeled as AR random processes. The matrices are computed for the case of Gaussian sources, for which the resulting separation would be asymptotically optimal (approaching the corresponding Cramér Rao bound (CRB) for best possible separation [10]). Note that (at least) two other approaches, similarly exhibiting asymptotic optimality for this problem, are known in the literature: A maximum likelihood—based approach [7] by Dégerine and Zaïdi, and a Gaussian mutual information—based approach [19] by Pham. However, so far only WASOBI appears to be computationally applicable to large-scale problems [25]. Note further, that the optimal solution for the case of known spectra of the sources was characterized by Pham and Garat in [18]. We shall revisit this characterization in the sequel.

We assume AR sources of known maximal order $p_{\max}$, with distinct sets of AR coefficients. The observations' correlation matrices take the structure of (1) where $\mathbf{R}_{\mathbf{s}}[\tau] = \mathrm{diag}\{r_1[\tau], r_2[\tau], \ldots, r_d[\tau]\}$ are the sources' correlation matrices (diagonal due to the spatial independence of the sources), such that $r_k[\tau]$ is the autocorrelation of $s_k[n]$ at lag τ. It was shown in [9] that (asymptotically) the set of $M = p_{\max} + 1$ symmetric estimated correlation matrices

$$\widehat{\mathbf{R}}_{\mathbf{x}}[\tau] = \frac{1}{2N}\sum_{n=1}^{N}\left(\mathbf{x}[n]\mathbf{x}^T[n+\tau] + \mathbf{x}[n+\tau]\mathbf{x}^T[n]\right) \quad (25)$$

at lags $\tau = 0, \ldots, M-1$ (assuming $N + M - 1$ samples are available) form a sufficient statistic for estimating $\mathbf{A}_0$.

To estimate the elements of $\mathrm{cov}[\widehat{\mathbf{r}}_x]$, we shall assume that the observations are (nearly) separated and are therefore (nearly) statistically independent. Throughout this section we shall use the simplified notations $\widehat{R}_{k\ell}[\tau]$ and $R_{k\ell}[\tau]$ instead of $(\widehat{\mathbf{R}}_{\mathbf{x}}[\tau])_{k\ell}$ and $(\mathbf{R}_{\mathbf{x}}[\tau])_{k\ell}$, respectively. Under the Gaussianity assumption we have [20]

$$\lim_{N\to\infty} N\,E\{(\widehat{R}_{k\ell}[\tau] - R_{k\ell}[\tau])(\widehat{R}_{pq}[\tau'] - R_{pq}[\tau'])\} = \delta_{kp}\delta_{\ell q}\xi_{k\ell}[\tau' - \tau] + \delta_{kq}\delta_{\ell p}\xi_{k\ell}[\tau' + \tau] \quad (26)$$

where $\delta_{k\ell}$ is the Kronecker delta and

$$\xi_{k\ell}[\tau] = \sum_{m=-\infty}^{\infty} R_{kk}[m]R_{\ell\ell}[\tau - m]. \quad (27)$$

Note that under the near-separation assumption $R_{kp}[m] = \delta_{kp}r_k[m]$ so that $\mathrm{cov}[\widehat{\mathbf{r}}_{k\ell}, \widehat{\mathbf{r}}_{pq}] = \mathbf{0}$ unless $k = p$ and $\ell = q$, which establishes the block-diagonality of $\mathrm{cov}[\widehat{\mathbf{r}}_{\mathbf{x}}]$ and of the associated (optimal) $\mathbf{W}$.

It can be shown that $\xi_{k\ell}[0], \ldots, \xi_{k\ell}[2M-2]$ can be computed as the correlation sequence of an AR process whose coefficients are given by a convolution of the kth and the ℓth AR coefficients. To see this, note that (27) can be rewritten as

$$\xi_{k\ell}[\tau] = \frac{1}{2\pi j}\oint S_k(z)S_\ell(z)z^{\tau-1}dz. \quad (28)$$

where $S_k(z)$ and $S_\ell(z)$ are Z-transforms of $r_k[\tau]$ and $r_\ell[\tau]$, respectively. Next, note that for an AR process with coefficients $a_{k0}, \ldots, a_{k,M-1}$ we have

$$S_k(z) = \frac{\sigma_k^2}{A_k(z)A_k^*\left(\frac{1}{z^*}\right)} \quad (29)$$

where $A_k(z) = \sum_{m=0}^{M-1} a_{km}z^{-m}$, and that the covariance sequence of each AR source is the inverse Z-transform of $S_k(z)$, namely

$$r_k[\tau] = \frac{1}{2\pi j}\oint S_k(z)z^{\tau-1}dz. \quad (30)$$

The aforementioned convolution relation follows from structural comparison of (27)–(30).

The computation of $\xi_{k\ell}[0], \ldots, \xi_{k\ell}[2M-2]$ can proceed by finding AR coefficients of processes with covariance functions $r_k[\tau]$ and $r_\ell[\tau]$, and by computing their convolution to form AR coefficients of an auxiliary AR process. Finally, $\xi_{k\ell}[0], \ldots, \xi_{k\ell}[2M-2]$ are found as the covariance function of the auxiliary process. The last step can be done in $O(M^2)$ operations, for example, by the inverse Schur and the inverse Levinson algorithms [20]. Furthermore, notice that the (p, q)th element of $\mathrm{cov}[\widehat{\mathbf{y}}_{k\ell}]$ reads

$$(\mathrm{cov}[\widehat{\mathbf{r}}_{k\ell}])_{pq} = (\xi_{k\ell}[p-q] + \xi_{k\ell}[p+q-2])(2 - \delta_{k\ell}) \quad (31)$$

therefore, $\mathrm{cov}[\widehat{\mathbf{r}}_{k\ell}]$ is a sum of Toeplitz and Hankel matrices, so its inverse $\mathbf{W}_{k\ell}$ can be computed in $O(M^2)$ operations, e.g., following the procedure in [12]. This procedure is used in the Matlab implementation of WASOBI [27].

When the sources' correlations $r_k[\tau]$ (or their AR parameters) are known, the resulting weight matrices are optimal (under

weak asymptotic conditions). When these optimal weight matrices are plugged into the "normal equations" in (20), a condition for the optimality of $\widehat{\mathbf{V}}$ is obtained. It is interesting to examine whether this optimality condition coincides with the optimality condition (for separation of stationary sources with known spectra) derived by Pham and Garat in [18], which, in our notations (for the problem at hand) can be expressed as

$$\sum_{\tau=-M}^{M} \phi_k[\tau](\widehat{\mathbf{V}}\widehat{\mathbf{R}}_{\mathbf{x}}[\tau]\widehat{\mathbf{V}}^T)_{k\ell} = 0 \quad 1 \le k \ne \ell \le d. \tag{32}$$

Here $\phi_k[\tau]$ are the Fourier coefficients of the inverse of the spectral density of the $k-$th source. The answer is, indeed, positive, and we provide a proof of this equivalence in Appendix D.

As an interesting byproduct, our proof shows that $\mathbf{W}_{k\ell}\mathbf{r}_{\ell\ell}$, the product of the optimum weight matrix and the vector of true covariances of the ℓth source, is independent of ℓ, and depends only on the spectrum of the kth source. Consequently, it is possible to design an algorithm which is similar to WASOBI, but does not need to compute all $d(d-1)/2$ weight matrices $\mathbf{W}_{k\ell}$, but only d sequences $\phi_k[\tau]$, $k = 1, \ldots, d$ instead. However, the presentation of such an algorithm exceeds the scope of this paper.

B. Block-Stationary-White Sources: BGL

The BGL algorithm [19] was proposed for BSS of block-wise stationary [independently, identically distributed (i.i.d.) in each block] Gaussian random processes. In this section, we propose an alternative approach, which is based on the iterative process outlined in Fig. 1, and is therefore asymptotically optimal, just like BGL.

The block-stationarity assumption in [19] asserts that the observation interval $n = 1, 2, \ldots, N$ can be divided into M nonoverlapping intervals, each of length N_m, in which each of the sources is Gaussian and i.i.d., having zero mean and variance $(\mathbf{R}_{\mathbf{s}}[m])_{ii} > 0$, where i denotes the source index and m denotes the interval index. Our proposed separation approach is based on the AJD of the set of the observations' covariance matrices from all intervals, $\widehat{\mathbf{R}}_{\mathbf{x}}[m]$, $m = 0, \ldots, M-1$, estimated using straightforward averaging within each interval.

It can be easily seen that if the sources are nearly separated, we have $\widehat{\mathbf{R}}_{\mathbf{x}}[m] \approx \mathbf{R}_{\mathbf{x}}[m] \approx \mathbf{R}_{\mathbf{s}}[m]$, and, since each source is assumed to be i.i.d. in time

$$\mathrm{var}(\widehat{R}_{k\ell}[m]) = \frac{(\mathbf{R}_{\mathbf{s}}[m])_{kk}(\mathbf{R}_{\mathbf{s}}[m])_{\ell\ell}}{N_m} \tag{33}$$

and $\widehat{R}_{k\ell}[m]$ and $\widehat{R}_{k'\ell'}[m']$ are uncorrelated unless $k = k'$, $\ell = \ell'$ and $m = m'$.

Therefore, the optimal weight matrix $\mathbf{W}$ is not only block-diagonal, but also diagonal in this case, and its $k\ell$th diagonal block satisfies

$$(\mathbf{W}_{k\ell})_{m+1,m+1} = \frac{N_m}{(\mathbf{R}_{\mathbf{s}}[m])_{kk}(\mathbf{R}_{\mathbf{s}}[m])_{\ell\ell}} \quad m = 0, 1, \ldots, M-1. \tag{34}$$

Like BGL, the iteratively reweighted WEDGE algorithm, in which the weight matrices are computed using the estimated $\widehat{R}_{kk}[m]$ in lieu of the true variances in (34), is asymptotically optimal. In practice, under asymptotic conditions, the two algorithms produce results which are almost identical.

C. Block-Stationary-AR Sources: Block-WASOBI

The data models introduced in the two previous subsections can be combined into one, allowing for block-wise stationary Gaussian sources modeled as different AR (rather than white) processes in each interval (and independent between intervals). The separation would be based on applying WEDGE to the set of $M_b \cdot M_d$ lagged covariance matrices, where M_b is the number of blocks and M_d is the number of delays (time lags), equal to the maximum assumed AR order plus one. Derivation of the optimum weight matrices is straightforward by combining the results of the previous two sections. Such a combined model might be useful for separating sources which can be modeled as block-wise stationary, being colored within blocks—such as speech signals.

The advantages of the resulting algorithm are twofold: Naturally, it offers (asymptotically) optimal exploitation of both the nonstationarity and the spectral diversity; But furthermore, assume that one wishes to only use zero-lagged correlations in each interval ($M_d = 1$, as in BGL): When the sources are not white within intervals, straightforward application of BGL to these matrices would no longer be (asymptotically) optimal, not even with respect to these matrices alone, since the implied weighting of BGL is equivalent to the weighting obtained under the whiteness assumption. With the proposed algorithm, proper weighting can be obtained, outperforming BGL in such cases.

VI. Simulations

We first present a comparative simulation study of typical convergence patterns and running speeds of U-WEDGE versus state-of-the-art competing generic AJD algorithms operating on synthetic large-scale target-matrices. Then, we demonstrate the performance of the weighted version (WEDGE) in the three BSS scenarios considered in Section V, and compare to competing algorithms in terms of accuracy and speed.

A. Generic AJD With Uniform Weights

1) Positive Definite Matrices: We generated $M = 10$ matrices of dimension $d \times d$ with $d = 20$, 100 as follows: $\mathbf{R}_{\mathbf{s}}[0]$ was always set to the identity matrix, whereas $\mathbf{R}_{\mathbf{s}}[m]$ (for $m = 1, \ldots, M-1$) were generated as diagonal matrices with all diagonal elements drawn independently from a Uniform distribution between 1 and 2, $\mathcal{U}(1,2)$. Then $\widehat{\mathbf{R}}_{\mathbf{x}}[m]$ were generated as $\widehat{\mathbf{R}}_{\mathbf{x}}[m] = \mathbf{A}_0\mathbf{R}_{\mathbf{s}}[m]\mathbf{A}_0^T + \sigma/2(\mathbf{N}[m] + \mathbf{N}^T[m])$, where $\mathbf{A}_0$ is the mixing matrix, $\mathbf{N}[k]$ is a noise matrix with i.i.d. elements drawn from a standard Normal distribution $\mathcal{N}(0,1)$, and σ is a free parameter. In this example, we chose $\mathbf{A}_0$ to be orthogonal, generated using the QR decomposition (see, e.g., [11]) of a random matrix with i.i.d. $\mathcal{N}(0,1)$ elements (taking the $\mathbf{Q}$ factor).

Fig. 2 presents typical convergence patterns of LLAJD [17], FFDIAG [33], QAJD [30], FAJD [16], QRJ2D [1] and our proposed U-WEDGE. The convergence is shown in terms of the unweighted criterion C_{LS} of (2) over a single trial. From our

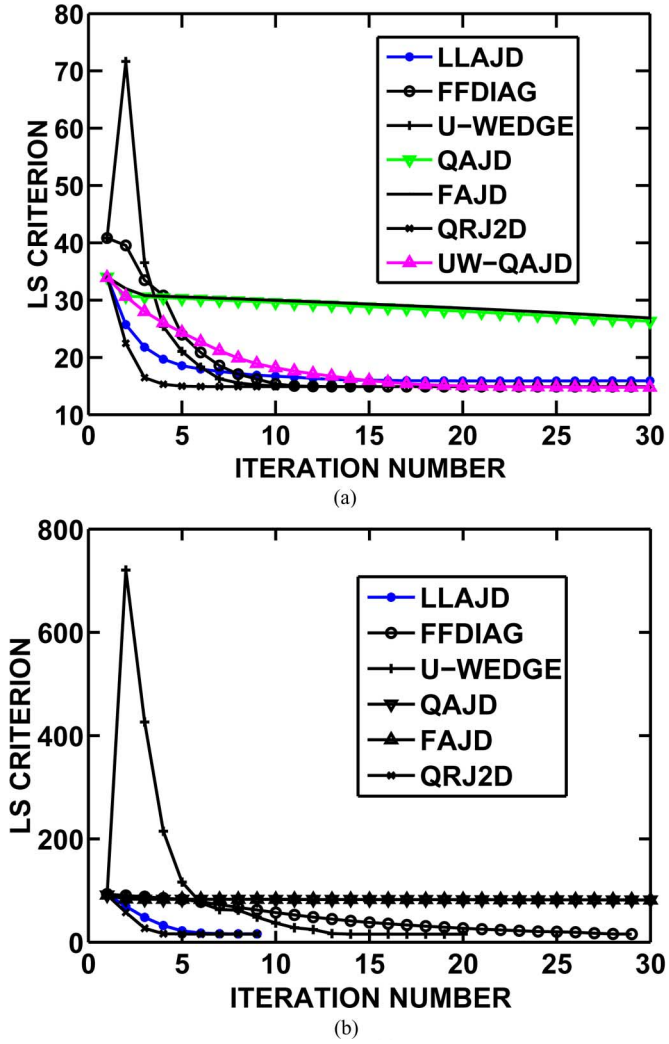

Fig. 2. The LS criterion (2) versus the iteration number for $\sigma = 0.1$, $d = 20$ in diagram (a) and $\sigma = 0.02$ and $d = 100$ in diagram (b).

experience, there is no considerable variation in these patterns between trials (when using the same matrix-generation setup). We note the tendency of U-WEDGE to increase C_{LS} in the first iteration, before decreasing rapidly in the succeeding iterations—note, however, that U-WEDGE was not designed to minimize C_{LS} (or any other specific criterion), so this behavior is not a particular anomaly. We note the significantly slower convergence (in terms of number of iterations) of QAJD and FAJD, especially with the higher matrix dimensions $(d = 100)$.

We observe that in the large-scale case $(d = 100)$, LLAJD and QRJ2D converged after approximately seven iterations, UWAJD after 15 and FFDIAG after 25 or 30. QAJD and FAJD needed 30 iterations or more. To complement this information, we present in Table I the actual running-times (on a Pentium 4 PC, 3.4–GHz with 1-GB RAM, running in Matlab 7.0.4 on Windows XP Professional) of the algorithms with the specified number of iterations. Here LLAJD stands for the original (Pham's) implementation of Pham's algorithm, whereas LLAJDp stands for a more efficient implementation thereof (In short, each iteration of LLAJD consists of a full sweep which requires $d(d-1)/2$ internal minimizations. In LLAJD, these

TABLE I
ASYMPTOTIC COMPUTATIONAL COMPLEXITIES OF AJD ALGORITHMS AND COMPUTATION TIMES IN MATLAB FOR $d = 100$ AND $M = 10$, POSITIVE DEFINITE MATRICES, ORTHOGONAL MIXING MATRIX

Algorithm	Complexity per iter.	No. of iter. used	Running time [s]
LLAJD	O(Md^3)	7	11
LLAJDp	O(Md^3)	7	3.5
FFDIAG	O(Md^3)	30	7.3
QAJD	O(Md^3)	30	56
FAJD	O($Md^3 + d^4$)	30	100
QRJ2D	O(Md^3)	7	12
U-WEDGE	O(Md^3)	**15**	**0.40**

internal minimizations are performed sequentially. In LLAJDp, they are partially parallelized, saving looping opeartions in Matlab).

Note that the theoretical asymptotic computational complexities (per iteration) of the algorithms are quite similar, as it is generally dominated by the complexity of the transformations $\widehat{\mathbf{R}}_{\mathbf{s}}[m] = \widehat{\mathbf{V}}\widehat{\mathbf{R}}_{\mathbf{x}}[m]\widehat{\mathbf{V}}^T$ for $m = 0, 1, \ldots, M-1$. However, the actual computation times are very different, because the different algorithms entail different programming structures in terms of vector and matrix operations in Matlab, possibly reducing sequential looping and enabling parallel processing.

2) Indefinite Matrices, Nonorthogonal Mixing: The previous experiment was modified such that the distribution of the diagonal elements of $\mathbf{R}_{\mathbf{s}}[m]$ (for $m = 1, \ldots, M-1$) was taken as $\mathcal{U}(-1, 1)$ (rather than $\mathcal{U}(1, 2)$), giving rise to sign-indefinite target-matrices. The mixing matrix $\mathbf{A}_0$ was taken at random with i.i.d. $\mathcal{N}(0, 1)$ elements and subsequently had its columns normalized such that each row of $\mathbf{V}_0 = \mathbf{A}_0^{-1}$ had unit Euclidean norm [15]. The results are shown in Fig. 3. We note that the convergence of QAJD and FAJD is not as (relatively) slow as in the previous example.

B. Weighted AJD in BSS

We now turn to present examples of applying the iteratively reweighted scheme (Fig. 1) to the three types of BSS problems presented in Section V.

1) WASOBI: We consider blind separation of $d = 100$ AR sources with distinct spectra. Such an example can serve to demonstrate suitability of the algorithm for separating high-dimensional data sets such as high density EEG signals.

The target-matrices $\widehat{\mathbf{R}}_{\mathbf{x}}[m]$ are sample covariance matrices of the mixture at lags τ, for $\tau = 0, \ldots, 10$. The AR processes had poles at $p_k^{(i)} e^{\pm j\pi k/6}$, $k = 1, \ldots, 5$ where $p_k^{(i)} \in \{0.6\rho, 0.85\rho, 0.95\rho\}$ and ρ is a free parameter. Among the $3^5 = 243$ possible distinct AR processes, which differ in the modulus of at least one pair of complex conjugate poles, 100 processes were chosen for the test. The parameter ρ allows to tune the spectral dynamic range of the sources. For small ρ, the sources' power spectra are flat, similar to each other and therefore hard to separate. The spectra become more distinct as 0.95ρ approaches 1. Sources of length $N = 16000$ were mixed by random matrices (with a condition number ≤ 5) in 100 independent trials.

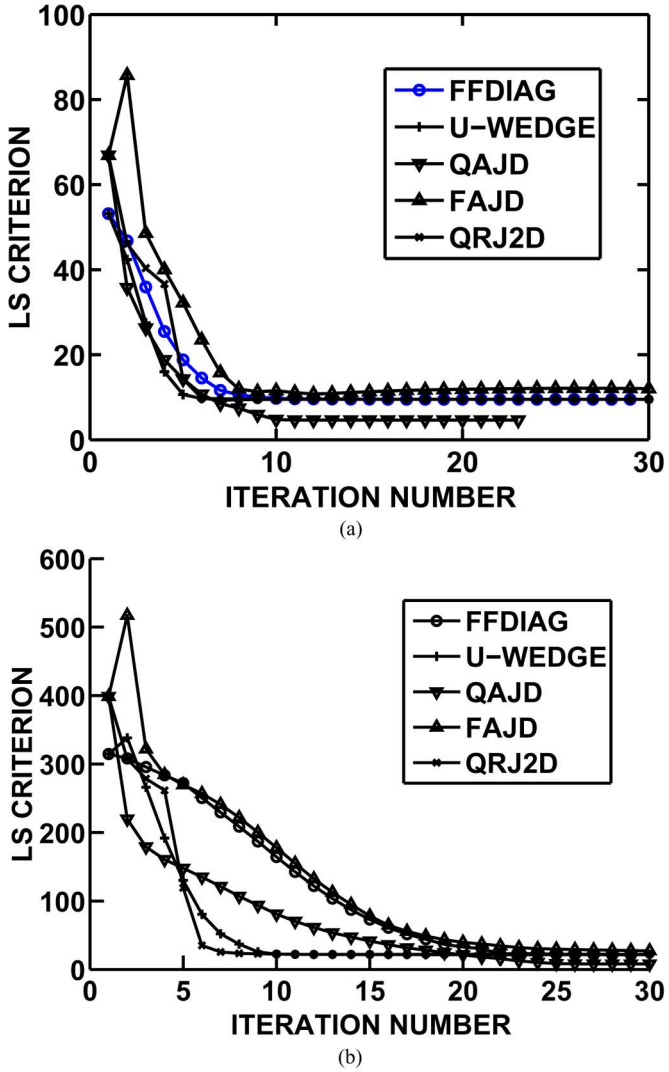

Fig. 3. The LS criterion (2) versus the iteration number for indefinite target matrices, $\sigma = 0.1$, $d = 20$, and $\sigma = 0.02$, and $d = 100$, respectively.

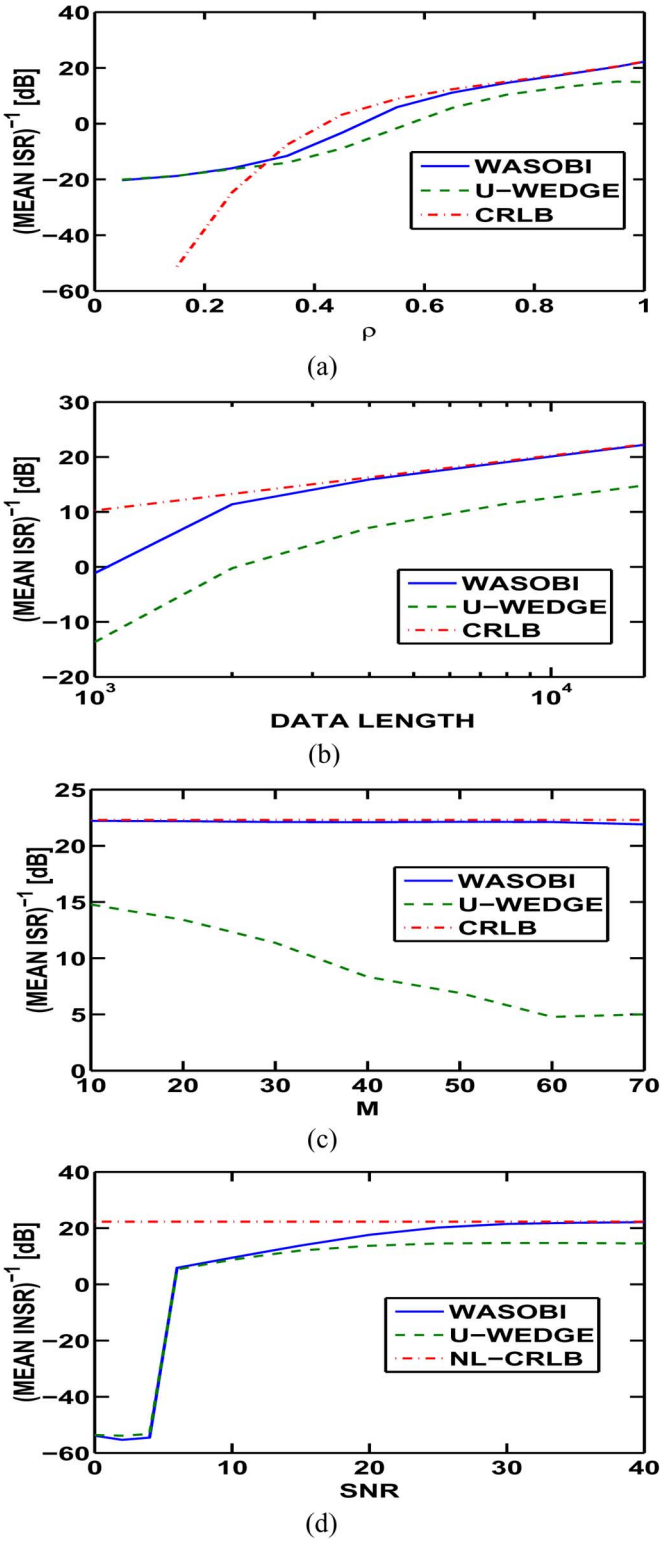

Fig. 4. Inverted mean ISR of 100 AR sources separated by WASOBI and U-WEDGE versus (a) parameter ρ, (b) the data length N, (c) the number of estimated correlation matrices LM, and (d) the input SNR (added noise). The CRB in (d) pertains to the noiseless scenario.

The ISR's of the algorithms FFDIAG and QAJD are not shown, as they are undistinguishable by naked eye from the results obtained by U-WEDGE. The average computation times were about 1s for the initial separation by U-WEDGE. Each application of WEDGE required about 3s and the entire procedure required 27s of CPU time.

In Fig. 4, diagrams (a)–(d), the inverted mean ISR and the corresponding ISR bound computed from the corresponding CRLB [10] are plotted versus N (the data length) for $\rho = 1$ and $M = 10$; versus ρ for $N = 16000$; and versus M (the number of estimated correlation matrices), respectively. The fourth diagram shows the performance in presence of additive Gaussian noise, with the CRLB computed for the noiseless case. Performance is also compared to that attained in the preprocessing stage by U-WEDGE.

We note the asymptotic efficiency of WASOBI, as well as its robustness with respect to overestimating the order of the AR processes (by using more correlation matrices), contrasted by the adverse impact on U-WEDGE. Comparison to SOBI in such a large-scale $(d = 100)$ example would take prohibitively long running time (for 100 trials), and the results are expected to be similar to those of U-WEDGE (since they are both use uniformly weighted AJD).

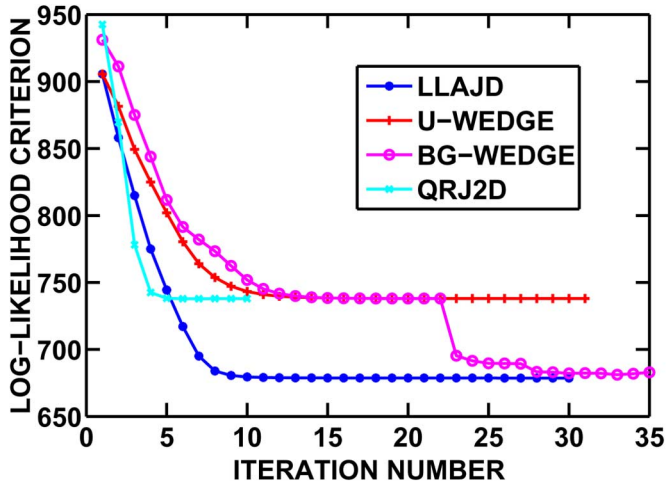

Fig. 5. The Pham's log-likelihood criterion (3) versus the iteration number for different AJD algorithms.

2) BGL: We now address the scenario of block-stationary sources, white within each block. The target-matrices are therefore sample covariance matrices (at zero lag), taken from $M_b = 40$ data segments of length 100 samples each. In each segment the $d = 20$ sources were generated as spectrally white Gaussian processes, with variances drawn at random (independently in each segment), from a $\mathcal{U}(0,1)$ distribution. The mixing matrix was the same orthogonal random matrix $\mathbf{A}_0$ as in the first experiment.

Our iteratively reweighted scheme of Fig. 1 was applied (using WEDGE) with the weights prescribed in Section V-B above. We used 20 iterations of the unweighted (U-WEDGE) phase, followed by three outer (reweighting) iterations, each consisting of at most five inner (WEDGE) iterations. We name the resulting algorithm BG-WEDGE.

The maximum likelihood (ML) estimator of the mixing/demixing matrices in this case is realized by Pham's LLAJD algorithm, and thus the most meaningful diagonalization criterion in this case is the log-likelihood criterion (3).

Fig. 5 shows typical learning curves for LLAJD (implemented using the efficient parallelized version, LLAJDp), QRJ2D, U-WEDGE and BG-WEDGE. As expected, we can see that BG-WEDGE achieves approximately the same performance as LLAJD, since both are asymptotically optimal (each with its own reasoning). The other algorithms, U-WEDGE and QRJ2D are only suboptimal. This is also apparent from the inverted average ISR's obtained by the four techniques in 100 independent trials. They were nearly the same, equal to 38.15 dB for LLAJD and BG-WEDGE and equal to 33.94 dB for both U-WEDGE and QRJ2D. In other words, separation performance of BG-WEDGE and LLAJD in terms of ISR was in average 4 dB better than that of the other two techniques.

Running times, excluding the computation of the covariance matrices (which is common to all algorithms), were 5.62s, 3.03s, 0.21s, and 0.29s for LLAJDp, QRJ2D, U-WEDGE and BG-WEDGE, respectively. Thus, the asymptotic optimality of BG-WEDGE in this example in Matlab is attained at a very low computational cost compared to LLAJD/LLAJDp.

3) Block WASOBI: We now address the framework of block-stationary sources which are not spectrally white within blocks (but are modeled as unknown AR processes). We compare the block-WASOBI algorithm (iteratively reweighted with weights prescribed in Section V-C) to that of WASOBI and BGL, which ignore the spectral shape of the sources within blocks. Rather than use synthetic sources, we chose to use real natural speech signals in this example, since such signals roughly fit the block-stationary model, but are certainly not white within blocks.

We used speech utterances of $d = 15$ different speakers (male and female), sampled at 8000 Hz. In each trial, each source was taken as $N = 6000$ samples of speech by the respective speakers, starting at random times. The sources were mixed using a random orthogonal matrix and blindly separated by the three methods. We applied some coarse optimization of the tuning variables of the algorithms. For WASOBI, we selected the AR order 12, for BGL and Block WASOBI we selected partitioning to 20 intervals of equal length. For Block WASOBI, the best separation was obtained for AR order equal to one in each interval. The resultant average signal-to-interference ratios after the demixing were 26.3 dB for WASOBI, 20.7 dB for BGL, and 30.5 dB for the Block WASOBI.

The experiment was then repeated using a nonorthogonal mixing matrix, generated as the sum of the Identity matrix and a random matrix with i.i.d. elements drawn from a $\mathcal{N}(0,1)$ distribution. In this way, speech energy of all speakers is not equally present in the mixtures. The resultant average SIR was thereby reduced to 21.2, 14.9, and 23.6 dB for WASOBI, BGL, and Block WASOBI, respectively.

These results clearly demonstrate the advantage of Block-WASOBI in separation of speech signals.

VII. Conclusion

We introduced a novel AJD algorithm, given the acronym WEDGE, with two clear advantages over competing AJD algorithms: significant computational efficiency, and the ability to accommodate weight matrices, which can considerably enhance (and even optimize) the separation performance in BSS. We also pointed out the relation to FFDIAG [33] and to QAJD [30], deriving a weighted version (W-QAJD) of the latter.

The theoretical asymptotic complexity of WEDGE and U-WEDGE is the same as that of most of its competitors, $O(Md^3)$. In Matlab, however, our implementation of U-WEDGE was shown to run significantly faster than all of its competitors. This is mainly due the convenient way in which U-WEDGE lends itself to be coded in terms of vector and matrix operations, avoiding looping iterations, which are rather time-consuming in Matlab.

For the computation of weight matrices we considered an iteratively reweighted scheme, and prescribed the estimation of asymptotically optimal weights in three scenarios involving Gaussian sources: Stationary AR sources, Block-stationary sources which are white within blocks, and Block-stationary sources which are colored (AR processes) within blocks. These three scenarios gave rise to the development of WASOBI, BG-WEDGE, and Block-WASOBI algorithms, respectively, all utilizing WEDGE for the weighted AJD and U-WEDGE for the initial, unweighted phase.

The resulting WASOBI algorithm is especially suited for separation in large-scale problems such as EEG and MEG. BG-WEDGE offers a computationally attractive alternative to BGL in the Block-stationary-white scenario, and can also be combined with WASOBI into Block-WASOBI, whose scope extends also to block-wise-colored sources such as speech signals.

The entire framework of AJD considered in this paper consisted of real-valued, symmetric target-matrices. When the target-matrices are complex-valued and Hermitian, and the diagonal source matrices are real-valued, extension of our algorithms for finding complex-valued mixing/demixing matrices is straightforward, and merely involves substitution of all "transpose" operations with "conjugate transpose" (thus, the same Matlab code can be readily applied in such cases). In the fully complex framework, in which the diagonal matrices may also be complex-valued (and the target-matrices non-Hermitian), adaptation of our algorithms is more involved, but possible, also [27].

Full Matlab code of the algorithms is available online at [27].

APPENDIX A
A PROPOSED WEIGHTED VERSION OF QAJD

In this Appendix, an implementation of the weighted QAJD, termed W-QAJD is presented. It is based on a slightly modified Newton's method. In the special case of uniform weights, the algorithm is called uniformly weighted QAJD (UW-QAJD) and serves as an alternative implementation of QAJD.[6] The development of the W-QAJD algorithm would provide some insight into the small-errors perturbations of the solution, for comparison with the WEDGE solution.

Assume that the algorithm operates near the optimum solution, i.e., assume that the constrained minimum of the criterion (18) is achieved for $\widehat{\mathbf{V}}$ being in a close neighborhood of the identity matrix. Then, the criterion and the constraint may be expressed in terms of $\widehat{\mathbf{V}}$ using Taylor series expansions at $\widehat{\mathbf{V}} = \mathbf{I}$. An approximate solution for $\widehat{\mathbf{V}}$ would then be found by minimizing a quadratic criterion under a linear constraint. The next step would be the same as in WEDGE: substitute $\{\widehat{\mathbf{R}}[m]\}$ with $\{\widehat{\mathbf{V}}\widehat{\mathbf{R}}[m]\widehat{\mathbf{V}}^T\}$, and iterate several times until convergence to a stationary point is reached.

More specifically, let $\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}})$ in (13) be expanded in terms of its first-order Taylor expansion as

$$\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}}) = \widehat{\mathbf{r}}^0_{k\ell} + \mathbf{G}_{k\ell}\widehat{\mathbf{v}} \tag{35}$$

where $\widehat{\mathbf{v}} = \mathrm{vec}(\widehat{\mathbf{V}} - \mathbf{I})$, $\widehat{\mathbf{r}}^0_{k\ell} = [(\widehat{\mathbf{R}}[0])_{k\ell}, \ldots, (\widehat{\mathbf{R}}[M-1])_{k\ell}]^T$, and $\mathbf{G}_{k\ell}$ is a suitable $M \times d^2$ matrix, namely

$$\mathbf{G}_{k\ell} = \begin{bmatrix} \mathbf{0} & (\widehat{\mathbf{R}}[0])_{k,:} & \mathbf{0} & (\widehat{\mathbf{R}}[0])_{\ell,:} & \mathbf{0} \\ & \vdots & & & \\ \mathbf{0} & (\widehat{\mathbf{R}}[M-1])_{k,:} & \mathbf{0} & (\widehat{\mathbf{R}}[M-1])_{\ell,:} & \mathbf{0} \end{bmatrix} \tag{36}$$

where the two nonzero $M \times d$ blocks in $\mathbf{G}_{k\ell}$ are at positions ℓ and k, respectively, for $\ell, k = 1, \ldots, d$, $\ell < k$.

[6]Unlike the original QAJD algorithm, UW-QAJD does not seem to exhibit significant differences in convergence rates between cases of positive-definite or sign-indefinite target matrices.

The simplified criterion is

$$C_{\mathrm{SIM}}(\widehat{\mathbf{v}}) = \sum_{k>\ell} \left(\widehat{\mathbf{r}}^0_{k\ell} + \mathbf{G}_{k\ell}\widehat{\mathbf{v}}\right)^T \mathbf{W}_{k\ell} \left(\widehat{\mathbf{r}}^0_{k\ell} + \mathbf{G}_{k\ell}\widehat{\mathbf{v}}\right). \tag{37}$$

The linearized constraint is given by $\mathrm{diag}((\widehat{\mathbf{V}} - \mathbf{I})\mathbf{R}[0]) = \mathbf{0}$ or $\mathbf{C}\widehat{\mathbf{v}} = \mathbf{0}$, where

$$\mathbf{C} = \begin{bmatrix} (\mathbf{R}[0])_{1,:} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & (\mathbf{R}[0])_{2,:} & \ldots & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & \ldots & (\mathbf{R}[0])_{d,:} \end{bmatrix}. \tag{38}$$

In other words, any admissible $\widehat{\mathbf{v}}$ must lie in an orthogonal complement of the columnspace of $\mathbf{C}^T$. Let a matrix $\mathbf{H}$ denote a basis of this orthogonal complement. In Matlab $\mathbf{H}$ can be found, e.g., by `qr` decomposition of $\mathbf{C}^T$. Then, any admissible $\widehat{\mathbf{v}}$ can be written as $\widehat{\mathbf{v}} = \mathbf{H}\widehat{\mathbf{w}}$, where $\widehat{\mathbf{w}}$ is any $(d^2 - d) \times 1$ vector. The constrained minimizer of the criterion (37) can then be found as the unconstrained minimum of $C_{\mathrm{SIM}}(\mathbf{H}\widehat{\mathbf{w}})$ with respect to $\widehat{\mathbf{w}}$. Straightforward computation yields the optimum $\widehat{\mathbf{v}} = \mathbf{H}\widehat{\mathbf{w}}$ as

$$\widehat{\mathbf{v}} = -\mathbf{H}\left[\mathbf{H}^T\left(\sum_{k>\ell}\mathbf{G}^T_{k\ell}\mathbf{W}_{k\ell}\mathbf{G}_{k\ell}\right)\mathbf{H}\right]^{-1} \cdot \mathbf{H}^T\left(\sum_{k>\ell}\mathbf{G}^T_{k\ell}\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{k\ell}\right). \tag{39}$$

Then, the $d^2 \times 1$ vector $\widehat{\mathbf{v}}$ is reshaped into a $d \times d$ matrix, to which the identity matrix would be added to yield the estimated $\widehat{\mathbf{V}}$. In practice, the convergence rate (in terms of the typical number of required iterations) appears to be as fast as that of the WEDGE. However, as we can see from (39), each iteration requires the solution of a linear system of dimension $(d^2 - 2) \times (d^2 - d)$, so that its overall computational complexity is $O(d^6)$. Therefore, this implementation is not suitable for large-dimensional matrices encountered in large-scale BSS.

The derivation of the W-QAJD allows, however, asymptotic analysis of the algorithm under the assumption that the target matrices differ from diagonal matrices by small perturbations, and $\widehat{\mathbf{R}}[0] \approx \mathbf{I}$. Then, the $d \times d^2$ matrix $\mathbf{C}$ is approximately a zeros-ones matrix, such that $\mathbf{C} \cdot \mathrm{vec}(\mathbf{X})$ consists only of diagonal elements $\mathbf{X}$ for any $d \times d$ matrix $\mathbf{X}$. Thus, an orthogonal complement $\mathbf{H}$ of such $\mathbf{C}^T$ approximately equals to a 0–1 matrix such that $\mathbf{H} \cdot \mathrm{vec}(\mathbf{X})$ is a $(d^2 - d)$ column vector composed of all off-diagonal elements of $\mathbf{X}$. It therefore follows that perturbation in the constraint matrix $\widehat{\mathbf{R}}[0]$ only influence the diagonal elements of $\widehat{\mathbf{V}}$, but not its off-diagonal elements. Further, under these conditions it can be seen that the off-diagonal elements of the optimum $\widehat{\mathbf{V}}$ can be found by solving the 2×2 systems

$$\begin{bmatrix} \hat{V}_{k\ell} \\ \hat{V}_{\ell k} \end{bmatrix} = -\begin{bmatrix} \widehat{\mathbf{r}}^T_{\ell\ell}\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{\ell\ell} & \widehat{\mathbf{r}}^T_{kk}\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{\ell\ell} \\ \widehat{\mathbf{r}}^T_{kk}\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{\ell\ell} & \widehat{\mathbf{r}}^T_{kk}\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{kk} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\mathbf{r}}^T_{\ell\ell}\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{k\ell} \\ \widehat{\mathbf{r}}^T_{kk}\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{k\ell} \end{bmatrix}. \tag{40}$$

Note that when $\widehat{\mathbf{V}}$ is close to identity, and $\widehat{\mathbf{A}} = \widehat{\mathbf{V}}^{-1}$, we have $\hat{V}_{k\ell} \approx -\hat{A}_{k\ell}$. Comparison of (24) and (40) reveals that WEDGE and W-QAJD have the same first-order perturbations under the condition that the target matrices $\widehat{\mathbf{R}}[m]$ are nearly diagonal.

APPENDIX B
RELATION BETWEEN $\boldsymbol{\Theta}(\mathbf{V})=\mathbf{I}$ AND $\boldsymbol{\Psi}(\mathbf{V})=\mathbf{0}$

A sufficient condition for $\boldsymbol{\Theta}(\mathbf{V})=\mathbf{I}$ is

$$\left.\frac{\partial \widetilde{C}_{LS}(\mathbf{V},\mathbf{A})}{\partial A_{k\ell}}\right|_{\mathbf{A}=\mathbf{I}}=0 \tag{41}$$

for $k,\ell = 1,\ldots,d$, where $\widetilde{C}_{LS}(\mathbf{V},\mathbf{A})$ was defined in (7). Let $B_{m,k,\ell}$ denote the (k,ℓ)th element of $\mathbf{V}\mathbf{R}_{\mathbf{x}}[m]\mathbf{V}^T$. Then straightforward calculus gives that the condition (41) is always satisfied for $k=\ell$. For $k\neq\ell$ we have

$$\begin{aligned}
&\left.\frac{\partial \widetilde{C}_{LS}(\mathbf{V},\mathbf{A})}{\partial A_{k\ell}}\right|_{\mathbf{A}=\mathbf{I}}\\
&=\frac{\partial}{\partial A_{k\ell}}\sum_{m,p,q}\left(B_{m,p,q}-\sum_r A_{pr}B_{m,r,r}A_{qr}\right)^2\Bigg|_{\mathbf{A}=\mathbf{I}}\\
&=-2\sum_{m,p,q}B_{m,p,q}\,\frac{\partial}{\partial A_{k\ell}}(A_{p\ell}B_{m,\ell,\ell}A_{q\ell})\Big|_{\mathbf{A}=\mathbf{I}}\\
&=-4\sum_{m=0}^{M-1}B_{m,k,\ell}B_{m,\ell,\ell}.
\end{aligned}\tag{42}$$

In matrix form, (41) is equivalent to $\boldsymbol{\Psi}(\mathbf{V})=\mathbf{0}$, where $\boldsymbol{\Psi}(\mathbf{V})$ was defined in (9).

APPENDIX C

This appendix presents a method that allows for any set of matrices $\mathbf{R}_{\mathbf{x}}[m]$, $m=1,\ldots,M-1$, to find two symmetric matrices $\mathbf{R}_{\mathbf{x}}[M]$ and $\mathbf{R}_{\mathbf{x}}[M+1]$ such that the augmented set of the matrices obeys

$$\operatorname{off}\left[\sum_{m=0}^{M+1}\mathbf{R}_{\mathbf{x}}[m]\cdot\operatorname{ddiag}(\mathbf{R}_{\mathbf{x}}[m])\right]=\mathbf{0}. \tag{43}$$

In this case, $\mathbf{A}=\mathbf{I}$ may not necessarily be argmin of $\widetilde{C}_{LS}(\mathbf{I},\mathbf{A})$ defined in (7), but it is a stationary point, where the gradient is zero.

Put

$$\mathbf{S}=-\left[\sum_{m=0}^{M-1}\mathbf{R}_{\mathbf{x}}[m]\cdot\operatorname{ddiag}(\mathbf{R}_{\mathbf{x}}[m])\right]. \tag{44}$$

The condition (43) is then equivalent to $\operatorname{off}\{\mathbf{R}_{\mathbf{x}}[M]\cdot\operatorname{ddiag}(\mathbf{R}_{\mathbf{x}}[M])+\mathbf{R}_{\mathbf{x}}[M+1]\cdot\operatorname{ddiag}(\mathbf{R}_{\mathbf{x}}[M+1])\}=\operatorname{off}[\mathbf{S}]$ or

$$\begin{bmatrix}(\mathbf{R}_{\mathbf{x}}[M])_{ii} & (\mathbf{R}_{\mathbf{x}}[M+1])_{ii}\\ (\mathbf{R}_{\mathbf{x}}[M])_{jj} & (\mathbf{R}_{\mathbf{x}}[M+1])_{jj}\end{bmatrix}\begin{bmatrix}(\mathbf{R}_{\mathbf{x}}[M])_{ij}\\ (\mathbf{R}_{\mathbf{x}}[M+1])_{ij}\end{bmatrix}=\begin{bmatrix}S_{ij}\\ S_{ji}\end{bmatrix} \tag{45}$$

for $i,j=1,\ldots,d, i\neq j$. If the diagonals of $\mathbf{R}_{\mathbf{x}}[M]$ and $\mathbf{R}_{\mathbf{x}}[M+1]$ are chosen in advance so the 2×2 matrix in (45) are regular for each $i\neq j$ (for example, one can choose $(\mathbf{R}_{\mathbf{x}}[M])_{ii}=i$ and $(\mathbf{R}_{\mathbf{x}}[M+1])_{ii}=d-i+1$ for $i=1,\ldots,d$), the off-diagonal elements of $\mathbf{R}_{\mathbf{x}}[M]$ and $\mathbf{R}_{\mathbf{x}}[M+1]$ can be found by solving (45).

The condition means that the augmented set is a stationary solution for U-WEDGE and FFDIAG: If the algorithms are applied to such a set of matrices, they stop without doing anything. The stationary solution might or might not be the desired joint diagonalizer. In practice, however, convergence of the algorithm to such a false solution was never observed.

APPENDIX D
EQUIVALENCE OF WEDGE-WASOBI FOR AR SOURCES TO PHAM AND GARAT'S CONDITION

Pham and Garat [18] have shown that for separating stationary Gaussian sources, the ML estimate $\widehat{\mathbf{V}}$ (which is asymptotically optimal) has to satisfy the following set of equations (end of p.1718 in [18], restated in our notations):

$$\sum_{\tau=-N}^{N}\phi_k[\tau](\widehat{\mathbf{V}}\widehat{\mathbf{R}}_{\mathbf{x}}[\tau]\widehat{\mathbf{V}}^T)_{k\ell}=0\quad 1\le k\neq\ell\le d \tag{46}$$

where $\widehat{\mathbf{R}}_{\mathbf{x}}[\tau]$ is given by (25) (for all $|\tau|<N$, N being the observation length), and where $\phi_k[\tau]$ is a sequence whose Z-transform is the inverse of $S_k(z)$ of (29) (which is the Z-transform of $r_k[\tau]$, the correlation sequence of the kth source).

In our case of AR sources it is straightforward to show that $\phi_k[\tau]$ is a finite, symmetric sequence of (maximal) length $2p_{\max}+1$ (given by the scaled correlation sequence of the respective AR coefficients). Moreover, due to the Z-transform relation, we have the deconvolution property

$$\sum_{\tau=-\infty}^{\infty}\phi_k[\tau]r_k[n-\tau]=\sum_{\tau=-p_{\max}}^{p_{\max}}\phi_k[\tau]r_k[n-\tau]=\delta[n] \tag{47}$$

where $\delta[n]$ denoted Kronecker's delta. Next, define the sequence

$$\tilde{\phi}_k[\tau]=\frac{1}{4}\begin{cases}\phi_k[0] & \tau=0\\ 2\phi_k[\tau] & \tau\neq 0\end{cases} \tag{48}$$

and the $(p_{\max}+1)\times 1$ vectors

$$\tilde{\boldsymbol{\phi}}_k=[\,\tilde{\phi}_k[0]\quad\tilde{\phi}_k[1]\quad\ldots\quad\tilde{\phi}_k[p_{\max}]\,]^T \tag{49}$$

$$\mathbf{r}_k=[\,r_k[0]\quad r_k[1]\quad\cdots\quad r_k[p_{\max}]\,]^T. \tag{50}$$

Note that in this case, (46) can be expressed as

$$\begin{aligned}
&\tilde{\boldsymbol{\phi}}_k^T\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}})=0\\
&\tilde{\boldsymbol{\phi}}_\ell^T\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}})=0,\quad 1\le\ell<k\le d,
\end{aligned}\tag{51}$$

where $\widehat{\mathbf{r}}_{k\ell}(\widehat{\mathbf{V}})$ is defined in (12).

Our basic claim is that $\mathbf{W}_{k\ell}\mathbf{r}_\ell=\tilde{\boldsymbol{\phi}}_k$, and, likewise, $\mathbf{W}_{k\ell}\mathbf{r}_k=\tilde{\boldsymbol{\phi}}_\ell$ (where $\mathbf{W}_{k\ell}$ is the respective block of the WASOBI optimal weight matrix), and, therefore, (46) is equivalent to our WEDGE-WASOBI condition (20).

To prove this, we would show that the inverse relation holds, namely that

$$\mathbf{r}_\ell=\mathbf{W}_{k\ell}^{-1}\tilde{\boldsymbol{\phi}}_k=\mathbf{C}_{k\ell}\tilde{\boldsymbol{\phi}}_k \tag{52}$$

where $\mathbf{C}_{k\ell}$, the inverse of the optimal $\mathbf{W}_{k\ell}$, is the covariance of the estimated $\hat{\mathbf{r}}_{k\ell}$, whose (p,q)th element is given by [recall (27) and (31)]

$$(\mathbf{C}_{k\ell})_{pq} = 2(\xi_{k\ell}[p-q] + \xi_{k\ell}[p+q]) \quad p,q = 0,1,\ldots,p_{\max} \tag{53}$$

with

$$\xi_{k,\ell}[\tau] = \sum_{m=-\infty}^{\infty} r_\ell[m] r_k[\tau - m]. \tag{54}$$

Let us compute the pth element of the product $\mathbf{C}_{k\ell}\tilde{\boldsymbol{\phi}}_k$:

$$\begin{aligned}
\left(\mathbf{C}_{k\ell}\tilde{\boldsymbol{\phi}}_k\right)_p
&= \sum_{q=0}^{p_{\max}} (\mathbf{C}_{k\ell})_{pq}\tilde{\phi}_k[q] \\
&= 2\sum_{q=0}^{p_{\max}} (\xi_{k\ell}[p-q] + \xi_{k\ell}[p+q])\tilde{\phi}_k[q] \\
&= 2\left(\sum_{q=0}^{p_{\max}} \xi_{k\ell}[p-q]\tilde{\phi}_k[q] + \sum_{q=0}^{p_{\max}} \xi_{k\ell}[p+q]\tilde{\phi}_k[q]\right) \\
&= 2\left(\sum_{q=0}^{p_{\max}} \xi_{k\ell}[p-q]\tilde{\phi}_k[q] + \sum_{q=-p_{\max}}^{0} \xi_{k\ell}[p-q]\tilde{\phi}_k[-q]\right) \\
&= 2\left(\sum_{q=0}^{p_{\max}} \xi_{k\ell}[p-q]\tilde{\phi}_k[q] + \sum_{q=-p_{\max}}^{0} \xi_{k\ell}[p-q]\tilde{\phi}_k[q]\right) \\
&= 2\sum_{q=-p_{\max}}^{p_{\max}} \tilde{\phi}_k[q]\xi_{k\ell}[p-q] + \tilde{\phi}_k[0]\xi_{k\ell}[p] \\
&= \sum_{q=-p_{\max}}^{p_{\max}} \phi_k[q]\xi_{k\ell}[p-q].
\end{aligned} \tag{55}$$

Substituting $\xi_{k\ell}$ we get

$$\begin{aligned}
(\mathbf{C}_{k\ell}\tilde{\boldsymbol{\phi}}_k)_p &= \sum_{q=-p_{\max}}^{p_{\max}} \phi_k[q] \sum_{m=-\infty}^{\infty} r_\ell[m] r_k[p-q-m] \\
&= \sum_{m=-\infty}^{\infty} r_\ell[m] \sum_{q=-p_{\max}}^{p_{\max}} \phi_k[q] r_k[p-m-q] \\
&= \sum_{m=-\infty}^{\infty} r_\ell[m]\delta[p-m] \\
&= r_\ell[p]
\end{aligned} \tag{56}$$

where we have used (47) for the transition before the last.

This established the proof that $\mathbf{r}_\ell = \mathbf{W}_{k\ell}^{-1}\tilde{\boldsymbol{\phi}}_k$ and therefore $\mathbf{W}_{k\ell}\mathbf{r}_\ell = \tilde{\boldsymbol{\phi}}_k$. With straightforward substitution of the indices it can also be shown that $\mathbf{W}_{k\ell}\mathbf{r}_k = \tilde{\boldsymbol{\phi}}_\ell$. Thus, when the sources' spectra and correlation are known, (20) can be expressed (for $1 \le \ell < k \le d$) as

$$\begin{aligned}
\mathbf{r}_\ell^T \mathbf{W}_{k\ell}\hat{\mathbf{r}}_{k\ell}(\mathbf{V}) = 0 &\Leftrightarrow \tilde{\boldsymbol{\phi}}_k^T \hat{\mathbf{r}}_{k\ell}(\mathbf{V}) = 0 \\
\mathbf{r}_k^T \mathbf{W}_{k\ell}\hat{\mathbf{r}}_{k\ell}(\mathbf{V}) = 0 &\Leftrightarrow \tilde{\boldsymbol{\phi}}_\ell^T \hat{\mathbf{r}}_{k\ell}(\mathbf{V}) = 0.
\end{aligned} \tag{57}$$

ACKNOWLEDGMENT

The authors acknowledge the assistance of J. Nielsen in some computer simulations used in earlier version of this paper, and of E. Doron and Dr. T. Marek in pointing out some important insights. They also thank Dr. X.-L. Li for pointing out [1] and [2].

REFERENCES

[1] B. Afsari, Charleston and J. Rosca, Eds. *et al.*, "Simple LU and QR based non-orthogonal matrix joint diagonalization," in *Proc. 6th Int. Conf. Independent Compon. Analysis and Blind Source Separ.*, Berlin Heidelberg, Germany, 2006, vol. LNCS 3889, pp. 17–24, ICA 2006.

[2] B. Afsari, "Sensitivity analysis for the problem of matrix joint diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1148–1171, 2008.

[3] 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, pp. 434–444, Feb. 1997.

[4] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 17, pp. 161–164, 1996.

[5] J.-F. Cardoso, "On the performance of orthogonal source separation algrithms," in *Proc. EUSIPCO'94*, 1994, pp. 776–779.

[6] S. Dégerine and E. Kane, "A comparative study of approximate joint diagonalization algorithms for blind source separation in presence of additive noise," *IEEE Trans. Signal Process.*, vol. 55, pt. 2, pp. 3022–3031, 2007.

[7] 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, pp. 1492–1512, 2004.

[8] E. Doron and A. Yeredor, C. G. Puntonet and A. Prieto, Eds., "Asymptotically optimal blind separation of parametric Gaussian sources," in *Proc. ICA 2004*, 2004, pp. 390–397.

[9] E. Doron, "Asymptotically optimal blind separation of parametric Gaussian sources," M.Sc., Tel-Aviv Univ., Tel Aviv, Israel, 2003.

[10] 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, pp. 417–420, Jun. 2007.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 2nd, Ed. Baltimore, MD: John Hopkins Univ. Press, 1989.

[12] G. Heinig, P. Jankowski, and K. Rost, "Fast inversion algorithms of Toeplitz-plus-Hankel matrices," in *Numerische Mathematik*. New York: Springer-Verlag, 1988, vol. 52, pp. 665–682.

[13] C. J. James and C. W. Hesse, "Independent component analysis for biomedical signals," *Physiolog. Measure.*, vol. 26, pp. R15–R39, 2005.

[14] M. Joho and H. Mathis, "Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation," in *Proc. SAM 2002*, Rosslyn, VA, Aug. 4–6, , pp. 273–277.

[15] Z. Koldovský and P. Tichavský, "Methods of fair comparison of performance of linear ica techniques in presence of additive noise," in *Proc. 2006 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP-2006)*, Toulouse, France, May 14–19, 2006, vol. 5, pp. 873–876.

[16] X.-L. Li and X. D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solutions," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1803–1814, May 2007.

[17] D.-T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136–1152, 2001.

[18] D.-T. Pham and P. Garat, "Blind separation of mixtures of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, pp. 1712–1725, Jul. 1997.

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

[20] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[21] L. Shoker, S. Sanei, and J. Chambers, "Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm," *IEEE Signal Process. Lett.*, vol. 12, pp. 721–724, Oct. 2005.

[22] H. W. Sorenson, *Parameter Estimation*. New York: Dekker, 1980.

[23] A. Smekhov and A. Yeredor, "Optimization of JADE using a novel optimally weighted joint diagonalization approach," in *Proc. 12th Eur. Conf. Signal Process. (EUSIPCO2004)*, Vienna, Austria, Sep. 2004.
[24] A. C. Tang, M. T. Sutherland, and C. J. McKinney, "Validation of SOBI components from high-density EEG," *Neuroimage*, vol. 25, no. 2, pp. 539–553, Apr. 2005.
[25] 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*, Florence, Italy, Sep. 4–8, 2006.
[26] P. Tichavsky, Z. Koldovsky, A. Yeredor, G. Gomez-Herrero, and E. Doron, "A hybrid technique for blind separation of non-Gaussian and time-correlated sources using a multicomponent approach," *IEEE Trans. Neural Netw.*, vol. 19, no. 3, pp. 421–430, Mar. 2008.
[27] P. Tichavský, Matlab Code for U-WEDGE, WEDGE, BG-WEDGE and WASOBI [Online]. Available: http://si.utia.cas.cz/Tichavsky.html
[28] P. Tichavský, A. Yeredor, and J. Nielsen, "A fast approximate joint diagonalization algorithm using a criterion with a block diagonal weight matrix," in *Proc. ICASSP 2008*, Las Vegas, NV, Mar.–Apr. 30-4, 2008, pp. 3321–3324.
[29] A.-J. van der Veen, "Joint diagonalization via subspace fitting techniques," in *Proc. IEEE ICASSP*, Salt Lake City, UT, May 2001.
[30] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3270–3278, Sep. 2006.
[31] A. Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *IEEE Signal Process. Lett.*, vol. 7, pp. 197–200, Jul. 2000.
[32] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, pp. 1545–1553, Jul. 2002.
[33] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777–800, 2004.

Petr Tichavský (M'98–SM'04) received the M.S. degree in mathematics in 1987 from the Czech Technical University, Prague, and the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences in 1992.

Since that time he has been with the Institute of Information Theory and Automation, Academy of Sciences , Czech Republic, Prague. In 1994, he received a Fulbright grant for a ten–month fellowship with Department of Electrical Engineering, Yale University, New Haven, CT. He is an author and coauthor of research papers in the area of sinusoidal frequency/frequency-rate estimation, adaptive filtering and tracking of time varying signal parameters, algorithm-independent bounds on achievable performance, sensor array processing, independent component analysis, and blind signal separation.

Dr. Tichavský received the Otto Wichterle Award from the Academy of Sciences of the Czech Republic in 2002. He served as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2002 to 2004, and since 2005, has served as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is a member of the IEEE Signal Processing Society's Signal Processing Theory and Methods (SPTM) Technical Committee.

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 he teaches courses 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 has been awarded the yearly Best Lecturer of the Faculty of Engineering Award five times. He is a member of the IEEE Signal Processing Society's Signal Processing Theory and Methods (SPTM) Technical Committee.