Blind Separation of Gaussian Sources via Second-Order Statistics with Asymptotically Optimal Weighting

Arie Yeredor, *Member, IEEE*

***Abstract*—Blind separation of Gaussian sources with different spectra can be attained using second-order statistics. The second-order blind identification (SOBI) algorithm, proposed by Belouchrani *et al.*, uses approximate joint diagonalization. We show that substantial improvement over SOBI can be attained when the joint diagonalization is transformed into a properly weighted nonlinear least squares problem. We provide an iterative solution and derive the optimal weights for our weights-adjusted SOBI (WASOBI) algorithm. The improvement is demonstrated by analysis and simulations.**

***Index Terms*—Blind source separation, joint diagonalization, weighted least squares.**

I. Introduction

BLIND source separation (BSS) addresses the reconstruction of N statistically independent source signals from M linear combinations thereof. In the static mixture framework, the observation model is $\boldsymbol{x}[t] = \boldsymbol{A}\boldsymbol{s}[t]$ $t = 1,2,\cdots T$, where $\boldsymbol{s}[t] = [s_1[t]s_2[t]\cdots s_N[t]]^T$ are the source signals, $\boldsymbol{x}[t] = [x_1[t]x_2[t]\cdots x_M[t]]^T$ are the observations, and $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ is the unknown mixing matrix. The term blind ascribes the lack of any additional information regarding the signals or $\boldsymbol{A}$.

When the signals are either nonstationary or stationary with different spectra, second-order statistics may be used to attain consistent estimates of $\boldsymbol{A}$. In [1], Belouchrani *et al.*proposed the second-order blind identification (SOBI) algorithm for stationary signals. The observations' correlation matrices $\boldsymbol{R}_x[\tau] \triangleq E[\boldsymbol{x}[t+\tau]\boldsymbol{x}^H[t]]$ satisfy

$$\boldsymbol{R}_x[\tau] = \boldsymbol{A}\boldsymbol{R}_s[\tau]\boldsymbol{A}^H \quad \forall\tau \tag{1}$$

where $\boldsymbol{R}_s[\tau] \triangleq E[\boldsymbol{s}[t+\tau]\boldsymbol{s}^H[t]]$ are the source signals' (unknown) diagonal correlation matrices. Thus, $\boldsymbol{A}$ is a joint diagonalizer of any set of K matrices $\{\boldsymbol{R}_x[\tau_1], \boldsymbol{R}_x[\tau_2], \cdots \boldsymbol{R}_x[\tau_K]\}$. In addition, it can be shown that if all the source signals have different spectra (differing by more than scale), then a set of lags can be found such that the joint diagonalizer is unique, up to irrelevant scaling and permutation of columns.

It is therefore proposed in [1] to estimate $\boldsymbol{A}$ as the joint diagonalizer of a set of estimated correlation matrices $\hat{\boldsymbol{R}}_x[\tau_1]$, $\hat{\boldsymbol{R}}_x[\tau_2]$, $\cdots \hat{\boldsymbol{R}}_x[\tau_1]$. However, while the set of true correlation matrices admits exact diagonalization, it is almost surely impossible to jointly diagonalize the set of estimated matrices. It is still possible, however, to obtain consistent estimators for $\boldsymbol{A}$ by resorting to approximate joint diagonalization, attained in [1] in two phases.

In the first phase, a whitening matrix $\hat{\boldsymbol{W}}$ is found such that $\hat{\boldsymbol{W}}\hat{\boldsymbol{R}}_x[0]\hat{\boldsymbol{W}}^H$ equals the identity matrix. All the other matrices are then similarly transformed $\hat{\tilde{\boldsymbol{R}}}[\tau_k] = \hat{\boldsymbol{W}}\hat{\boldsymbol{R}}_x[\tau_k]\hat{\boldsymbol{W}}^H$ $k = 1,2,\cdots K$. In the second phase, the unitary approximate joint diagonalizer $\hat{\boldsymbol{U}}$ of the transformed set is found using successive Jacobi rotations, which iteratively minimize the off-diagonal entries of the transformed matrices (see [1], [3]). The desired estimate is then given by $\boldsymbol{A} = \hat{\boldsymbol{W}}^{\#}\hat{\boldsymbol{U}}$ where $\hat{\boldsymbol{W}}^{\#}$ denotes the pseudo-inverse of $\hat{\boldsymbol{W}}$).

It can be easily observed (see also [5]) that the second phase optimizes a least-squares (LS) fit of the $\hat{\tilde{\boldsymbol{R}}}[\tau_k]$-s with respect to $\hat{\boldsymbol{U}}$. However, this LS criterion is not optimized with respect to $\hat{\boldsymbol{A}}$, since the nonunitary part $\hat{\boldsymbol{W}}$ is chosen to attain *exact* diagonalization of $\hat{\boldsymbol{R}}_x[0]$, possibly at the expense of poor diagonalization of the other matrices. As noted in [2], such a "hard-whitening" operation bounds the attainable performance. Furthermore, the errors in estimating the correlations are strongly correlated. It is therefore expected that an unweighted LS criterion might yield inferior performance relative to an optimally weighted LS (WLS) criterion.

In this letter, we address these two shortcomings of the SOBI algorithm. First, we reformulate the approximate diagonalization problem as a nonlinear WLS problem and outline an iterative algorithm for minimization with respect to an arbitrary (not necessarily unitary) matrix $\hat{\boldsymbol{A}}$. We then find the optimal weight[1] matrix under the assumption of Gaussian source signals with finite-length correlations such as moving average (MA) processes. We demonstrate via error analysis (supported by simulations results) substantial improvement over the SOBI algorithm.

To capture the essence of our proposal in this limited-length exposition, we focus on the case of $M = N = 2$ real-valued signals with a real-valued mixing matrix. Extension to complex signals (and mixing) is relatively straightforward. For more than two signals, the algorithm can either be extended to higher dimensions at the cost of increased complexity, or be applied in couples at the cost of possibly degraded performance. Our algorithm is called weights-adjusted SOBI (WASOBI).

Manuscript received September 29, 1999. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. A. Nehorai.

The author is with the Department of Electrical Engineering Systems, Tel-Aviv University, Tel-Aviv, Israel (arie@eng.tau.ac.il).

Publisher Item Identifier S 1070-9908(00)05872-7.

[1]The weighting approach has also been proposed in [4], but not pursued further.

II. Formulation as a Weighted LS Problem

We assume that $\hat{\boldsymbol{R}}_x[\tau_k]$ are estimated using

$$\hat{\boldsymbol{R}}_x[\tau_k] = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}[t]\boldsymbol{x}^T[t+\tau_k] \tag{2}$$

(assuming $T+\tau_k$ samples are available).

We seek a 2×2 matrix and $\boldsymbol{A}$ and K diagonal matrices $\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2 \cdots \boldsymbol{\Lambda}_K$ such that $\hat{\boldsymbol{R}}_x[\tau_k]$ are "best fitted" by $\boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T$ for $k=1,2,\cdots K$. Thus, there are four parameters of interest, denoted $\boldsymbol{a} \triangleq \text{vec}\ \{\boldsymbol{A}\} = [\boldsymbol{A}^{(1,1)}\boldsymbol{A}^{(2,1)}\boldsymbol{A}^{(1,2)}\boldsymbol{A}^{(2,2)}]^T$, and $2K$ nuisance parameters, which are the K 2×1 vectors $\boldsymbol{\lambda}_k \triangleq \text{diag}\ \{\boldsymbol{\Lambda}_k\}\ k=1,2,\cdots K$. However, due to the inherent scaling ambiguity (which enables one to commute scales between $\boldsymbol{A}$ and $\boldsymbol{\Lambda}_k$), we may arbitrarily fix, for example, $\boldsymbol{\Lambda}_1$, reducing the true number of nuisance parameters to $2(K-1)$.

Note that the estimated $\boldsymbol{R}_x[\tau_k]$ are not necessarily symmetric (for $\tau_k \neq 0$) in contrast to $\boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T$. We shall thus attempt to fit each $\boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T$ to a symmetric variant of the respective $\hat{\boldsymbol{R}}_x[\tau_k]$, obtained by substituting its off-diagonal terms with their arithmetic average. We therefore define $\hat{\boldsymbol{r}}_k \triangleq \text{vec}\ \{\hat{\boldsymbol{R}}_x[\tau_k]\}$ and

$$\boldsymbol{y}_k \triangleq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \hat{\boldsymbol{r}}_k \triangleq \boldsymbol{C}\hat{\boldsymbol{r}}_k \qquad k=1,2\cdots K. \tag{3}$$

The desired fit for each k can then be written as

$$\boldsymbol{y}_k \approx \begin{bmatrix} a_1^2 & a_3^2 \\ a_1a_2 & a_3a_4 \\ a_2^2 & a_4^2 \end{bmatrix} \cdot \boldsymbol{\lambda}_k \triangleq \boldsymbol{G}(\boldsymbol{a})\boldsymbol{\lambda}_k. \tag{4}$$

Concatenating all $\boldsymbol{y}_k$ into $\boldsymbol{y} \triangleq [\boldsymbol{y}_1^T\boldsymbol{y}_2^T\cdots\boldsymbol{y}_K^T]^T$, we get $\boldsymbol{y} \approx [\boldsymbol{I}_K \otimes \boldsymbol{G}(\boldsymbol{a})]\boldsymbol{\lambda} \triangleq \tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda}$, where $\boldsymbol{I}_K$ denotes the $K\times K$ identity matrix, $\otimes$ denotes Kronecker's product, and $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1^T\boldsymbol{\lambda}_2^T\cdots\boldsymbol{\lambda}_K^T]^T$ is the concatenation of $\boldsymbol{\lambda}_k$. We also define $\overline{\boldsymbol{\lambda}} = [\boldsymbol{\lambda}_2^T\boldsymbol{\lambda}_3^T\cdots\boldsymbol{\lambda}_K^T]^T$, the vector of free parameters in $\boldsymbol{\lambda}$. Given any $3K\times 3K$ symmetric weight matrix $\boldsymbol{W}$, we may now define the WLS criterion as

$$C_{WLS}(\boldsymbol{a},\boldsymbol{\lambda}) \triangleq [\boldsymbol{y}-\tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda}]^T\boldsymbol{W}[\boldsymbol{y}-\tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda}] \tag{5}$$

to be minimized with respect to $\boldsymbol{a}$ and $\overline{\boldsymbol{\lambda}}$, with $\boldsymbol{\lambda}_1$, set arbitrarily. While linear in $\overline{\boldsymbol{\lambda}}$, this WLS criterion is nonlinear in $\boldsymbol{a}$. To use Gauss' method (e.g., [6]), we differentiate $\tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda}$ with respect to $\boldsymbol{a}$, obtaining

$$\boldsymbol{H}(\boldsymbol{a},\boldsymbol{\lambda}) \triangleq \frac{\partial(\tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda})}{\partial \boldsymbol{a}} = \begin{bmatrix} \lambda_1^{(1)}\boldsymbol{D}_1(\boldsymbol{a}) & \lambda_1^{(2)}\boldsymbol{D}_2(\boldsymbol{a}) \\ \lambda_2^{(1)}\boldsymbol{D}_1(\boldsymbol{a}) & \lambda_2^{(2)}\boldsymbol{D}_2(\boldsymbol{a}) \\ \vdots & \vdots \\ \lambda_K^{(1)}\boldsymbol{D}_1(\boldsymbol{a}) & \lambda_K^{(2)}\boldsymbol{D}_2(\boldsymbol{a}) \end{bmatrix} \tag{6}$$

where $\boldsymbol{\lambda}_k^{(i)}$ denotes the ith element of $\boldsymbol{\lambda}_k$, and where

$$\boldsymbol{D}_1(\boldsymbol{a}) \triangleq \begin{bmatrix} 2a_1 & 0 \\ a_2 & a_1 \\ 0 & 2a_2 \end{bmatrix} \quad \boldsymbol{D}_2(\boldsymbol{a}) \triangleq \begin{bmatrix} 2a_3 & 0 \\ a_4 & a_3 \\ 0 & 2a_4 \end{bmatrix}. \tag{7}$$

The derivative of $\tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda}$ with respect to $\overline{\boldsymbol{\lambda}}$ is $\overline{\boldsymbol{G}}(\boldsymbol{a}) \triangleq \tilde{\boldsymbol{G}}(\boldsymbol{a})[\boldsymbol{0}\vdots\boldsymbol{I}_{2(K-1)}]^T$, where $\boldsymbol{0}$ denotes a $2(K-1)\times 2$ all-zeros matrix. The iterative Gauss algorithm thus assumes the following form:

$$\begin{bmatrix} \boldsymbol{a}^{[l+1]} \\ \overline{\boldsymbol{\lambda}}^{[l+1]} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}^{[l]} \\ \overline{\boldsymbol{\lambda}}^{[l]} \end{bmatrix} + \begin{bmatrix} \boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H} & \boldsymbol{H}^T\boldsymbol{W}\overline{\boldsymbol{G}} \\ \overline{\boldsymbol{G}}^T\boldsymbol{W}\boldsymbol{H} & \overline{\boldsymbol{G}}^T\boldsymbol{W}\overline{\boldsymbol{G}} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \boldsymbol{H}^T\boldsymbol{W} \\ \overline{\boldsymbol{G}}^T\boldsymbol{W} \end{bmatrix} [\boldsymbol{y}-\tilde{\boldsymbol{G}}\boldsymbol{\lambda}^{[l]}] \qquad l=1,2,\cdots \tag{8}$$

where $\boldsymbol{H}$, $\overline{\boldsymbol{G}}$, and $\tilde{\boldsymbol{G}}$ are shorthand for $\boldsymbol{H}(\boldsymbol{a}^{[l]},\boldsymbol{\lambda}^{[l]})$, $\overline{\boldsymbol{G}}(\boldsymbol{a}^{[l]})$, and $\tilde{\boldsymbol{G}}(\boldsymbol{a}^{[l]})$, respectively. An intelligent initial guess for $\boldsymbol{a}^{[0]}$ and $\boldsymbol{\lambda}^{[0]}$ can be obtained by using the SOBI algorithm.

The apparent computational load involved in applying (8) can be alleviated by exploiting the sparse structure of $\tilde{\boldsymbol{G}}(\boldsymbol{a})$ and also by alternating between linear minimization (with respect to $\boldsymbol{\lambda}$) and nonlinear minimization (with respect to $\boldsymbol{a}$). However, these computational aspects are beyond the scope of this letter. Note only that the computational load of the minimization depends only on K and is independent of the number of observations T.

III. Optimal Weighting

To apply optimal weighting, we need the covariance matrix of $\boldsymbol{y}$, denoted $\boldsymbol{\Phi}$. Assuming Gaussian signals, we have from (2)

$$\begin{aligned} &E[\hat{\boldsymbol{R}}_x^{(i,j)}[\tau_k]\hat{\boldsymbol{R}}_x^{(m,n)}[\tau_l]] \\ &\quad = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E[x_i[t]x_j[t+\tau_k]x_m[s]x_n[s+\tau_l]] \\ &\quad = \boldsymbol{R}_x^{(i,j)}[\tau_k]\boldsymbol{R}_x^{(m,n)}[\tau_l] \\ &\qquad + \frac{1}{T}\sum_{p=-(T-1)}^{T-1}\left(1-\frac{|p|}{T}\right) \\ &\qquad \cdot (\boldsymbol{R}_x^{(i,m)}[p]\boldsymbol{R}_x^{(j,n)}[p+\tau_l-\tau_k] \\ &\qquad + \boldsymbol{R}_x^{(i,n)}[p+\tau_l]\boldsymbol{R}_x^{(j,m)}[p-\tau_k]) \end{aligned} \tag{9}$$

which implies that the covariance of $\hat{\boldsymbol{R}}_x^{(i,j)}[\tau_k]$ and $\hat{\boldsymbol{R}}_x^{(m,n)}[\tau_l]$ is given by the expression of the last three rows. We now further assume that the source signals are MA processes of orders $\leq Q$, whereas the selected lags are $\tau_k = k-1, k=1,2,\cdots Q+1$. The summation over p can then be reduced from $-Q$ to Q for $1\leq k,l\leq K = Q+1$, which implies that estimating the correlation matrices up to lag Q is also sufficient for consistently estimating $\boldsymbol{\Phi}$.

Observe also that (9) can be reformulated in matrix form, such that

$$\begin{aligned} \text{Cov}[\hat{\boldsymbol{r}}_k,\hat{\boldsymbol{r}}_l] = \frac{1}{T}\sum_{p=-Q}^{Q}\left(1-\frac{|q|}{T}\right)&(\boldsymbol{R}_x[p+\tau_l-\tau_k]\otimes\boldsymbol{R}_x[p] \\ &+ (\boldsymbol{R}_x[p-\tau_k]\otimes\boldsymbol{R}_x[p-\tau_l])\boldsymbol{P}) \end{aligned} \tag{10}$$

where $\boldsymbol{P}$ is a permutation matrix that swaps the second and third columns of the matrix to its left. Recalling the linear transformation (4) from $\hat{\boldsymbol{r}}_k$ to $\boldsymbol{y}_k$, we conclude that the (k,l)th 3×3 block of $\boldsymbol{\Phi}$ is given by $\boldsymbol{\Phi}_{k,l} \triangleq \text{Cov}[\boldsymbol{y}_k,\boldsymbol{y}_l] = \boldsymbol{C}\ \text{Cov}[\hat{\boldsymbol{r}}_k,\hat{\boldsymbol{r}}_l]\boldsymbol{C}^T$. The optimal weight matrix is then given by $\boldsymbol{W}_{opt} = \boldsymbol{\Phi}^{-1}$. In

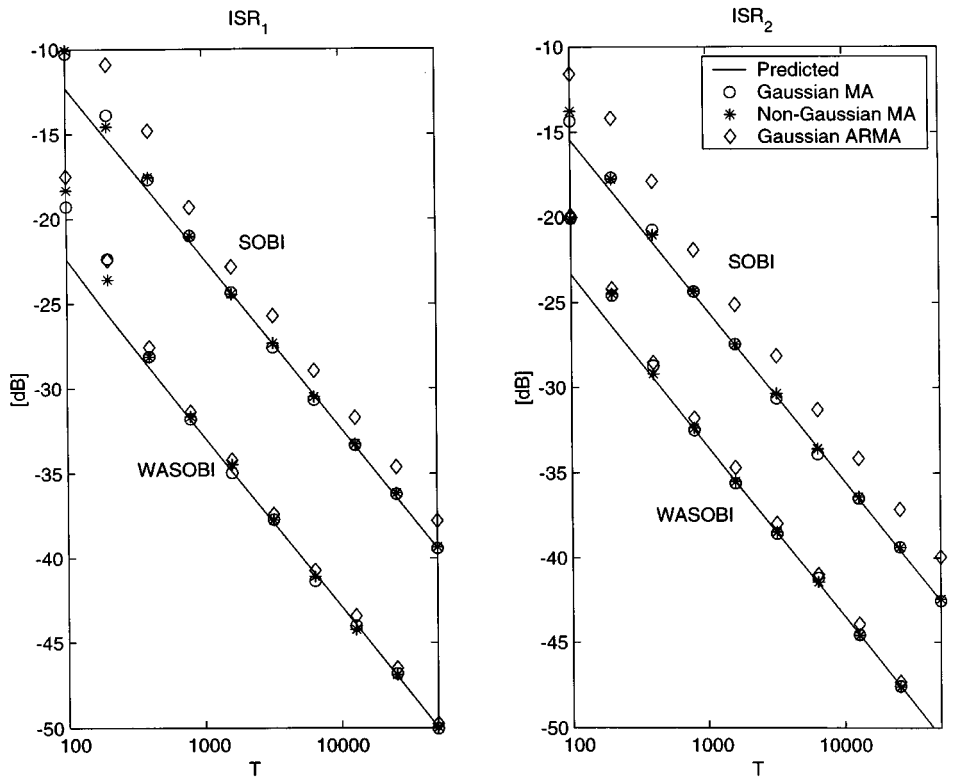

Fig. 1. Predicted ISR's (solid lines) and simulation results for SOBI and WASOBI. In the nominal setup, $s_1[t]$ is an MA(4) process with zeros at $0.8e^{\pm j(\pi/2)}$ and $0.7e^{\pm j(\pi/3)}$ (and their reciprocals). $s_2[t]$ is an MA(3) process with zeros at 0.7 and $0.3e^{\pm j(\pi/2)}$ (and their reciprocals). Results for Gaussian signals are denoted by "o," and for non-Gaussian signals (using uniformly-distributed driving noise), by "*." To demonstrate performance with non-MA signals (denoted by "◇"), auto-regressive moving average (ARMA) signals were generated by adding to each source signal a weak AR(1) signal with a pole at -0.95. Only lags 0 to 4 were used in each case. Improvement of about 10 dB of WASOBI over SOBI is evident in all cases. Both algorithms used the same data. Each simulation point represents an average of 1000 trials.

practice, estimated correlations would replace true correlations in (10), providing a consistent estimate of $\boldsymbol{W}_{opt}$.

IV. Performance Analysis and Simulations Results

The minimizers of (5) are denoted $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{\lambda}} = [\hat{\boldsymbol{\lambda}}_1^T \hat{\overline{\lambda}}^T]^T$, whereas $\boldsymbol{a}$ denotes the true entries of $\boldsymbol{A}$, and $\boldsymbol{\Lambda}_k = \boldsymbol{R}_s[\tau_k]$ denote the true source signals correlations. We assume temporarily that $\hat{\boldsymbol{\lambda}}_1$ is set to its true value $\boldsymbol{\lambda}_1 = \text{diag}\{\boldsymbol{R}_s[0]\}$, although this value is unknown (discussion will follow). As a result, there is no scaling ambiguity in $\hat{\boldsymbol{a}}$. Assuming small errors, $\hat{\boldsymbol{a}}$ and $\hat{\overline{\lambda}}$ are unbiased, and their joint covariance for any $\boldsymbol{W}$ is given by (e.g. [6])

$$\text{Cov}\begin{bmatrix}\hat{\boldsymbol{a}}\\ \hat{\overline{\lambda}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H} & \boldsymbol{H}^T\boldsymbol{W}\overline{\boldsymbol{G}}\\ \overline{\boldsymbol{G}}^T\boldsymbol{W}\boldsymbol{H} & \overline{\boldsymbol{G}}^T\boldsymbol{W}\overline{\boldsymbol{G}}\end{bmatrix}^{-1} \cdot \begin{bmatrix}\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{W}\boldsymbol{H} & \dot{\boldsymbol{H}}^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{W}\overline{\boldsymbol{G}}\\ \overline{\boldsymbol{G}}^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{W}\boldsymbol{H} & \overline{\boldsymbol{G}}^T\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{W}\overline{\boldsymbol{G}}\end{bmatrix} \cdot \begin{bmatrix}\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H} & \boldsymbol{H}^T\boldsymbol{W}\overline{\boldsymbol{G}}\\ \overline{\boldsymbol{G}}^T\boldsymbol{W}\boldsymbol{H} & \overline{\boldsymbol{G}}^T\boldsymbol{W}\overline{\boldsymbol{G}}\end{bmatrix}^{-1} \quad (11)$$

where $\boldsymbol{H}$ and $\overline{\boldsymbol{G}}$ are shorthand for $\boldsymbol{H}(\boldsymbol{a},\boldsymbol{\lambda})$ and $\overline{\boldsymbol{G}}(\boldsymbol{a})$, respectively. When $\boldsymbol{W}_{opt}$ is used, this expression reduces to its first term. Cov$[\hat{\boldsymbol{a}}]$ is given by the 4×4 upper-left block of (11).

Once $\hat{\boldsymbol{A}}$ is constructed from $\hat{\boldsymbol{a}}$, $\hat{\boldsymbol{A}}^{-1}$ can be applied to $\boldsymbol{x}(t)$ to retrieve $\boldsymbol{s}(t)$. To translate Cov$[\hat{\boldsymbol{a}}]$ into the residual interference to signal ratio (ISR), we seek the off-diagonal elements of $\hat{\boldsymbol{A}}^{-1}\boldsymbol{A}$, given by $(-a_4\hat{a}_3 + a_3\hat{a}_4)/|\hat{\boldsymbol{A}}|$ and $(a_2\hat{a}_1 - a_1\hat{a}_2)/|\hat{\boldsymbol{A}}|$ for the (1,2) and (2,1) elements, respectively. Under the small errors assumption, these terms have zero means with variances given by

$$\text{ISR}_1 \approx \frac{1}{|\boldsymbol{A}|^2}[0\ 0\ -a_4\ a_3]\ \text{Cov}\ [\hat{\boldsymbol{a}}][0\ 0\ -a_4\ a_3]^T$$
$$\text{ISR}_2 \approx \frac{1}{|\boldsymbol{A}|^2}[a_2\ -a_1\ 0\ 0]\ \text{Cov}\ [\hat{\boldsymbol{a}}][a_2\ -a_1\ 0\ 0]^T. \quad (12)$$

Although we assumed that $\boldsymbol{\Lambda}_1$ was set to its true (unknown) value of $\boldsymbol{R}_s[0]$, the ISR expressions in (12) hold valid even when $\boldsymbol{\Lambda}_1$ is set arbitrarily. This is because any (nonzero) setting for $\boldsymbol{\Lambda}_1$ merely inflicts scaling on the columns of $\hat{\boldsymbol{A}}$, translated to the rows of $\hat{\boldsymbol{A}}^{-1}\boldsymbol{A}$. While the retrieved signals are thus scaled, the resulting ISR's remain unchanged.

Equations (11) and (12) can be used to predict the performance with any $\boldsymbol{W}$. For example, for the SOBI (unweighted) algorithm, $\boldsymbol{W}$ would be set to $\boldsymbol{W}_{\text{SOBI}} = \text{diag}\ \{\alpha\boldsymbol{I}_3, \boldsymbol{I}_{3(K-1)}\}$ (in the matrices-to-matrix sense), where $\alpha \gg Q$ (we used α =100) is a large constant reflecting SOBI's obligatory whitening phase, which attributes infinite weight to warrant exact diagonalization of $\hat{\boldsymbol{R}}_x[0]$.

For WASOBI, $\boldsymbol{W}_{opt}$ is used. Note however, that since in practice $\boldsymbol{W}_{opt}$ is also estimated from the data, the predicted ISR's are only expected to be approached asymptotically, when the errors in estimating $\boldsymbol{W}_{opt}$ are negligible. In simulations, we used the estimated $\boldsymbol{W}_{opt}$.

Fig. 1 presents the predicted ISR's versus simulation results for both SOBI and WASOBI. Predictions are approached as T

increases, when the small-errors assumption prevails. The signals were MA(4) and MA(3) processes. We used[2] $\boldsymbol{A} = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$. It is evident that WASOBI significantly outperforms SOBI for all T's in this example.

To demonstrate (empirically) the robustness of WASOBI with respect to the MA and Gaussianity assumptions, we also provide simulation results for the cases of non-Gaussian signals and of infinite correlation auto-regressive moving average (ARMA) signals (using only lags 0 to 4).

[2]Interestingly, it turns out that neither the predicted nor the simulated performance (in terms of ISR's) depend on $\boldsymbol{A}$ for *neither* SOBI *nor* WASOBI. The $\boldsymbol{A}$-invariance of SOBI agrees with [1]. For WASOBI, it is more subtle to conclude from the derivation (invariance is not attained with arbitrary $\boldsymbol{W}$).

REFERENCES

[1] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.

[2] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," *Proc. 7th Eur. Signal Processing Conf. '94*, pp. 776–779, 1994.

[3] ——, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Applicat.*, vol. 17, no. 1, pp. 161–164, 1996.

[4] H. Sahlin, "Blind Signal Separation by Second Order Statistics," Ph.D. Thesis, School of Elect. Comp. Eng., Chalmers Univ. Technol., Goteborg, Sweden, 1998.

[5] M. Wax and J. Sheinvald, "A least-squares approach to joint diagonalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 52–53, Feb. 1997.

[6] H. W. Sorenson, *Parameter Estimation*. New York: Marcel-Dekker, 1980.