14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A COMPUTATIONALLY AFFORDABLE IMPLEMENTATION OF AN ASYMPTOTICALLY OPTIMAL BSS ALGORITHM FOR AR SOURCES

Petr Tichavský[1], *Eran Doron*[2], *Arie Yeredor*[2] *and Jan Nielsen*[1]

[1]Inst. of Information Theory and Automation,
Pod vodárenskou věží 4, P.O.Box 18,
182 08 Prague 8, Czech Republic
tichavsk@utia.cas.cz, nielsen@kaktus.cz

[2]School of Electrical Engineering
Tel-Aviv University
P.O.Box 39040, Tel-Aviv 69978, ISRAEL
eran@amimon.com, arie@eng.tau.ac.il

ABSTRACT

The second-order blind identification (SOBI) algorithm for separation of stationary sources was proved to be useful in many biomedical applications. This paper revisits the weights-adjusted variant of SOBI, known as WASOBI, which is asymptotically optimal (in separating Gaussian parametric processes), yet prohibitively computationally demanding for more than 2-3 sources. A computationally feasible implementation of the algorithm is proposed, which has a complexity not much higher than SOBI. Excluding the estimation of the correlation matrices, the post-processing complexity of SOBI is $O(d^4M)$, where d is the number of the signal components and M is the number of covariance matrices involved. The additional complexity of our proposed implementation of WASOBI is $O(d^6 + d^3M^3)$ operations. However, for WASOBI, the number M of the matrices can be significantly lower than that of SOBI without compromising performance. WASOBI is shown to significantly outperform SOBI in simulation, and can be applied, e.g., in the processing of low density EEG signals.

1. INTRODUCTION

The second-order blind identification (SOBI) algorithm is a classical blind source separation (BSS) algorithm for wide-sense stationary (WSS) processes with distinct spectra [1]. This algorithm has proved to be very useful in biomedical applications and became more popular than other algorithms [2, 3, 4].

This paper is focused on the optimal achievable separation. Hence we primarily study the SOBI algorithm that utilizes the full number of $O(d^4M)$ operations (excluding the $O(d^2MN)$ operations necessary for estimating the correlation matrices), where d is the number of sources, M is the number of correlation matrices used and N is the data length. In the literature, several extensions and modifications of the SOBI algorithm have been proposed - e.g., "robust SOBI" [7], "thin ICA" [6]. These algorithms are different in the sense that they do not perform all $O(d^4M)$ operations but a lower number, which allows e.g. processing of high density EEG signals with dimension 64 and more. The lower complexity, however, is attained at the cost of compromised separation accuracy, as will be shown in the simulation section.

This work was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572.

In [8], the SOBI algorithm was reformulated as a weighted nonlinear least squares (LS) problem, for which the optimum weights are in general different than those applied in SOBI. The use of asymptotically optimal weights gave rise to the weights-adjusted SOBI (WASOBI), which is asymptotically optimal (approaching the corresponding Cramér Rao bound (CRB) for best possible separation) for Gaussian moving average (MA) sources, [8], and later for sources modeled as Gaussian autoregressive (AR) and ARMA processes [9].

The estimation of the optimal weight matrix for the WASOBI algorithm requires the estimation of the observed processes parameters. Generally the observations are ARMA(p,q) processes, where $p = \sum_{i=1}^{d} p_i$, $q = max\{q_1+p-p_1, q_2+p-p_2, \ldots, q_d+p-p_d\}$ and p_i and q_i are i-th source's AR and MA orders, respectively (d is the number of sources). Then, the minimum number of estimated correlation matrices required to attain the asymptotic optimality equals $q + p + 1$, a number that increases linearly with the number of sources. Therefore, the direct implementation of WASOBI for AR or ARMA sources was highly computationally demanding and in practice it allowed blind separation of only 2 or 3 sources.

We concentrate the discussion in this paper on separation of AR sources for at least two reasons. First, AR processes are the most frequently used to model biomedical signals, MA or ARMA models are rarely used. Second, Gaussian AR processes are known to be the maximum entropy processes among all random processes having a prescribed set of values of covariance function. In this sense, the AR assumption is the least informative and hence the least committed to a particular structure.

It was shown that, when all the sources are Gaussian AR processes, a sufficient statistic for their separation is the set of estimated correlation matrix from lags 0 to their maximal AR-order [10], cf. Theorem 6.4 in [12]. In such cases, the estimation of correlation matrices for farther lags cannot improve the performance of an optimal estimation procedure.

In this paper, the WASOBI algorithm is re-visited and a novel implementation for AR sources is proposed. The new implementation involves about the same order of operations as SOBI (in our simulations, it needs about three times more operations than SOBI). At the present time it allows us to separate up to 20 sources of AR order 10, in computation time of about 2 minutes. The proposed approach is based on an iterative

scheme, in which a pre-processing separation algorithm (e.g., SOBI) is applied first, and then the optimal WASOBI is applied to the estimated correlation matrices of the nearly-separated sources. Consequently, the minimal number of estimated correlation matrices required for the estimation of the optimal weighting is reduced to the maximum assumed AR order plus one.

In simulations we show that the new implementation of WASOBI remains asymptotically efficient in the sense that it achieves the corresponding CRB in separating up to 20 sources, and may represent a vital alternative to SOBI in biomedical applications, for example in processing of low density EEG. In addition, we explore, in simulation, the robustness of WASOBI to additive noise (with which WASOBI is no longer claimed to be asymptotically optimal).

2. DATA MODEL AND METHOD

We address the noiseless static invertible model:

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] \qquad n = 1, 2, \ldots, N \qquad (1)$$

where $\mathbf{A} \in \mathbb{R}^{d \times d}$ is the unknown mixing matrix, $\mathbf{s}[n] = [s_1[n], \ldots, s_d[n]]^T \in \mathbb{R}^d$ are the unobserved statistically independent sources and $\mathbf{x}[n] = [x_1[n], \ldots, x_d[n]]^T \in \mathbb{R}^d$ are the d observed static linear mixtures of $\mathbf{s}[n]$. In our case, we further assume that the sources are Gaussian AR processes of known orders. The BSS problem consists of recovering the unobserved sources, or estimating the unknown mixtures coefficients.

The performance of BSS algorithms is usually measured by the interference to signal ratio (ISR), or its inverse signal-to-noise ratio (SIR).[1] In [10] the minimization of the ISRs was proven to be approximately equivalent to the minimization of the mean square error (MSE) in estimating the mixing matrix (assuming that the inherent scaling and permutation ambiguity are resolved). Therefore, by reformulating the problem as a weighted LS (WLS) problem, two goals are achieved: minimizing the MSE of the estimated $\mathbf{A}$ and data reduction: rather than estimate $\mathbf{A}$ from N vectors we estimate $\mathbf{A}$ from a small number of estimated correlation matrices. Such a data reduction is permissible since, as was proven in [10], cf. [12], the statistic:

$$\frac{1}{2}\left(\widehat{\mathbf{R}}_{\mathbf{x}}[\tau] + \widehat{\mathbf{R}}_{\mathbf{x}}^T[\tau]\right) \qquad \tau = 0, \ldots, p_{max} \qquad (2)$$

forms asymptotically sufficient statistic for the separation of AR Gaussian sources, where p_{max} is the maximal AR-order of the sources. We therefore choose $M = p_{max} + 1$ and compute

$$\widehat{\mathbf{R}}_{\mathbf{x}}[\tau] = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}[n]\mathbf{x}^T[n+\tau] \qquad \tau = 0, \ldots, M-1 \quad (3)$$

(assuming $N + M - 1$ samples are available). The observations' true correlation matrices take the structure

$$\mathbf{R}_{\mathbf{x}}[\tau] = \mathbf{A}\mathbf{R}_{\mathbf{s}}[\tau]\mathbf{A}^T \qquad \forall\tau \qquad (4)$$

[1] In the presence of additive noise, a preferred measure of performance is the signal-to-interference plus noise ratio (SINR) [11].

where due to the spatial independence of the sources, their correlation matrices $\mathbf{R}_{\mathbf{s}}[\tau] = \mathrm{diag}[\lambda_\tau^{(1)}, \lambda_\tau^{(2)}, \cdots, \lambda_\tau^{(d)}]$ are diagonal matrices and $\lambda_\tau^{(k)}$ is the auto-correlation of $s_k[n]$ at lag τ. This relation can also be written as

$$\mathrm{vec}\{\mathbf{R}_{\mathbf{x}}[\tau]\} = (\mathbf{A} \odot \mathbf{A})\boldsymbol{\lambda}_\tau \qquad (5)$$

where $\odot$ denotes the Khatri-Rao product (a column-wise Kronecker product), and $\boldsymbol{\lambda}_\tau = \mathrm{diag}\{\mathbf{R}_{\mathbf{s}}[\tau]\} = [\lambda_\tau^{(1)}, \lambda_\tau^{(2)}, \cdots, \lambda_\tau^{(d)}]^T$. Since the matrix $\mathbf{R}_{\mathbf{x}}[\tau]$ is symmetric, (5) can be further equivalently re-written as

$$\mathrm{svec}\{\mathbf{R}_{\mathbf{x}}[\tau]\} = \mathbf{Q}(\mathbf{A} \odot \mathbf{A})\boldsymbol{\lambda}_\tau \qquad (6)$$

where $\mathrm{svec}\{\mathbf{R}\}$ is similar to $\mathrm{vec}\{\mathbf{R}\}$, but only stacks the $\frac{d(d+1)}{2}$ elements of $\frac{1}{2}(\mathbf{R} + \mathbf{R}^T)$ lying on and below its main diagonal, and $\mathbf{Q}$ is a suitable matrix composed of zeros, ones, and $\frac{1}{2}$'s.

Introducing the notation $\mathbf{y}_\tau \triangleq \mathrm{svec}\{\mathbf{R}_x[\tau]\}$, and $\mathbf{G}(\mathbf{A}) \triangleq \mathbf{Q}(\mathbf{A} \odot \mathbf{A})$, (6) can be written as

$$\mathbf{y}_\tau = \mathbf{G}(\mathbf{A})\boldsymbol{\lambda}_\tau \ . \qquad (7)$$

Given M estimated correlation matrices $\widehat{\mathbf{R}}_{\mathbf{x}}[\tau]$, transformed as $\widehat{\mathbf{y}}_\tau = \mathrm{svec}\{\widehat{\mathbf{R}}_{\mathbf{x}}[\tau]\}$, the overall LS formulation of the problem is:

$$\widehat{\mathbf{y}} \approx [\mathbf{I}_M \otimes \mathbf{G}(\mathbf{A})]\boldsymbol{\lambda} \triangleq \mathbf{G}_o(\mathbf{A})\boldsymbol{\lambda}, \qquad (8)$$

where $\widehat{\mathbf{y}} = \left[\widehat{\mathbf{y}}_0^T, \widehat{\mathbf{y}}_1^T, \cdots, \widehat{\mathbf{y}}_{M-1}^T\right]^T$ is the new measurements vector, $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_0^T, \boldsymbol{\lambda}_1^T, \cdots, \boldsymbol{\lambda}_{M-1}^T]^T$.

According to the WLS approach, we have to minimize the following criterion:

$$C_{WLS}(\mathbf{A}, \boldsymbol{\lambda}) = [\widehat{\mathbf{y}} - \mathbf{G}_o(\mathbf{A})\boldsymbol{\lambda}]^T\mathbf{W}[\widehat{\mathbf{y}} - \mathbf{G}_o(\mathbf{A})\boldsymbol{\lambda}] \qquad (9)$$

with respect to $\mathbf{A}$, and $\boldsymbol{\lambda}$, where $\mathbf{W}$ is a weight matrix. The optimum weight matrix, which minimizes the mean square errors of the estimated parameters, is $\mathbf{W} = \{\mathrm{cov}[\widehat{\mathbf{y}}]\}^{-1}$. In the original WASOBI algorithm it was proposed to estimate the full matrix $\mathbf{W}$ from the estimated correlation matrices, basing the result on the relations between second- and fourth-order moments of the Gaussian sources. The resulting estimation procedure, followed by the tedious high-dimensional minimization procedure, resulted in a computationally prohibitive algorithm for cases of more than 2-3 sources.

The key observation which forms the basis for our proposed implementation, is that when the sources are nearly unmixed and the elements of $\widehat{\mathbf{y}}$ are properly reordered, together with the rows in $\mathbf{G}_o(\mathbf{A})$, the reordered optimal weight matrix, denoted $\widetilde{\mathbf{W}}$, becomes essentially block-diagonal, and therefore its computation from the estimated correlation becomes simpler. Moreover, the high-dimensional LS problem (9) can be decomposed into smaller individual LS problems - which offers substantial computational relief. Thus, the optimum weight matrix is estimated from correlation matrices of the nearly-separated sources, following an initial

separation attained by any closed form algorithm (e.g. SOBI). Iterating the loop (separation) - (estimation of $\widetilde{\mathbf{W}}$) - (updated separation) by minimizing (9) usually converges within up to three iterations, resulting in the same asymptotic efficiency (attaining the CRB) as original WASOBI.

Note further, that it is not necessary to actually separate the sources in the preprocessing (SOBI) stage and then re-estimate their correlations, since the estimated separating matrix can be applied directly to the original observations' estimated correlation matrices, yielding in turn the estimated correlations of the nearly-separated sources. The same holds true for each subsequent iteration of WASOBI.

The exact closed form expression for estimation of elements of $\widetilde{\mathbf{W}}$ will be shown later. However, we observe that $\widetilde{\mathbf{W}}$ is block diagonal. To be more specific, criterion (9) can then be rewritten in the form

$$C_{WLS}(\mathbf{A},\boldsymbol{\lambda}) = [\widetilde{\mathbf{y}} - \widetilde{\mathbf{G}}_o(\mathbf{A})\boldsymbol{\lambda}]^T \widetilde{\mathbf{W}} [\widetilde{\mathbf{y}} - \widetilde{\mathbf{G}}_o(\mathbf{A})\boldsymbol{\lambda}] \quad (10)$$

where

$$\widetilde{\mathbf{y}} = [\widetilde{\mathbf{y}}_{11}^T, \widetilde{\mathbf{y}}_{21}^T, \ldots, \widetilde{\mathbf{y}}_{22}^T, \widetilde{\mathbf{y}}_{32}^T, \ldots, \widetilde{\mathbf{y}}_{dd}^T]^T \quad (11)$$

$$\widetilde{\mathbf{G}}_o(\mathbf{A}) = [\mathbf{G}_{11}^T(\mathbf{A}), \mathbf{G}_{21}^T(\mathbf{A}), \ldots, \mathbf{G}_{dd}^T(\mathbf{A})^T]^T \quad (12)$$

$$\widetilde{\mathbf{W}} = \text{block diag}[\mathbf{W}_{11}, \mathbf{W}_{21}, \ldots, \mathbf{W}_{dd}]. \quad (13)$$

The τ-th element ($\tau = 1, \ldots, M$) of $\widetilde{\mathbf{y}}_{k\ell}$ is

$$(\widetilde{\mathbf{y}}_{k\ell})_\tau = \frac{1}{2}\left[\widehat{\mathbf{R}}_{k\ell}[\tau-1] + \widehat{\mathbf{R}}_{\ell k}[\tau-1]\right] \quad (14)$$

$$\mathbf{W}_{k\ell} = [\text{cov}(\widetilde{\mathbf{y}}_{k\ell})]^{-1} \quad (15)$$

$$\mathbf{G}_{k\ell}(\mathbf{A}) = \mathbf{I}_M \otimes (\mathbf{A}_{k\cdot} \star \mathbf{A}_{\ell\cdot}) \quad (16)$$

for $k, \ell = 1, \ldots d$, $k \geq \ell$, where $\mathbf{A}_{k\cdot}$ is the $k-$th row of $\mathbf{A}$ and "$\star$" is the elementwise product.

Exploiting the block-diagonal structure of $\widetilde{\mathbf{W}}$, the criterion $C_{WLS}(\mathbf{A},\boldsymbol{\lambda})$ in (10) can be simplified as

$$= \sum_{k,\ell,k\geq\ell} [\widetilde{\mathbf{y}}_{k\ell} - \widetilde{\mathbf{G}}_{k\ell}(\mathbf{A})\boldsymbol{\lambda}]^T \mathbf{W}_{k\ell} [\widetilde{\mathbf{y}}_{k\ell} - \widetilde{\mathbf{G}}_{k\ell}(\mathbf{A})\boldsymbol{\lambda}] \quad (17)$$

2.1 Computing elements of the weight matrix

For Gaussian sources it can be easily shown that

$$\begin{aligned}
&\lim_{N\to\infty} N\,\mathrm{E}\{(\widehat{\mathbf{R}}_{k\ell}[\tau] - \mathbf{R}_{k\ell}[\tau])(\widehat{\mathbf{R}}_{pq}[\tau'] - \mathbf{R}_{pq}[\tau'])\} \\
&= \lim_{N\to\infty} \sum_{m=-N+1}^{N-1} (1 - \frac{m}{N})(\mathbf{R}_{kp}[m]\mathbf{R}_{\ell q}[m+\tau'-\tau] \\
&\qquad + \mathbf{R}_{kq}[j+\tau']\mathbf{R}_{\ell p}[j-\tau]) \\
&\approx \phi^{(k,p)(\ell,q)}[\tau'-\tau] + \phi^{(k,q)(\ell,p)}[\tau'+\tau] \qquad (18)
\end{aligned}$$

where

$$\begin{aligned}
\phi^{(k,p)(\ell,q)}[\tau] &= \sum_{m=-\infty}^{\infty} \mathbf{R}_{kp}[m]\mathbf{R}_{\ell q}[m+\tau] \\
&= \frac{1}{2\pi j} \oint S^{(kp)}(z) S^{(\ell q)}(z) z^{\tau-1} dz \quad (19)
\end{aligned}$$

and $S^{(kp)}(z)$ is the cross-spectrum (in the Z-domain) between the k-th and p-th sources [9].

Note that for independent (separated) sources, $\mathbf{R}_{kp}[m] = 0$ unless $k = p$, so that $\phi^{(k,p)(\ell,q)}[\tau] = 0$ unless $k = p$ and $\ell = q$. The computational relief of the new implementation of WASOBI consists in part in the observation that in a neighborhood of the point of the separation, the values of $\phi^{(k,p)(\ell,q)}[\tau]$ for $k \neq p$ or $\ell \neq q$ need not be computed and can be neglected. This implies the block-diagonality of $\widetilde{\mathbf{W}}$.

Second, note that each integral in (19) can be computed in $O(M^2)$ operations as a variance of an AR process which has its coefficients given as a convolution of AR coefficients of the $k-$th and $\ell-$th signal [10].

2.2 Minimization of the weighted criterion

The main idea of the minimization of (10) is to write it as minimization of

$$C(\boldsymbol{\theta}) = [\widetilde{\mathbf{y}} - \mathbf{f}(\boldsymbol{\theta})]^T \widetilde{\mathbf{W}} [\widetilde{\mathbf{y}} - \mathbf{f}(\boldsymbol{\theta})] \quad (20)$$

where $\boldsymbol{\theta}$ is the unknown parameter composed of elements of $\mathbf{A}$ and $\boldsymbol{\lambda}$, and $\mathbf{f}(\boldsymbol{\theta}) = \mathbf{G}_o(\mathbf{A})\boldsymbol{\lambda}$. Then, a Gauss iterative method can be applied [13],

$$\boldsymbol{\theta}^{[i+1]} = \boldsymbol{\theta}^{[i]} + [\mathbf{F}_i^T \widetilde{\mathbf{W}} \mathbf{F}_i]^{-1} \widetilde{\mathbf{W}} \mathbf{F}_i [\widetilde{\mathbf{y}} - \mathbf{f}(\boldsymbol{\theta}^{[i]})] \quad (21)$$

where $\mathbf{F}_i = \partial \mathbf{f}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[i]}}$.

Note that the minimization problem (10) has the inherent BSS scaling ambiguity, which allows us to commute scales between the elements of $\boldsymbol{\lambda}$ and the columns of $\mathbf{A}$. Hence it is possible, without any loss in generality, to fix e.g., the first d elements of $\boldsymbol{\lambda}$ to arbitrary nonzero values (e.g. 1's), and exclude them from the minimization. Thus, the unknown parameters are $\boldsymbol{\theta} = [\boldsymbol{\theta}_{\mathbf{A}}^T, \overline{\boldsymbol{\lambda}}^T]^T$, where $\boldsymbol{\theta}_{\mathbf{A}} = \text{vec}\{\mathbf{A}^T\}$ and $\overline{\boldsymbol{\lambda}} = [\boldsymbol{\lambda}_1^T \boldsymbol{\lambda}_2^T \cdots \boldsymbol{\lambda}_{M-1}^T]^T$.

The iteration (21) can be rewritten as

$$\begin{aligned}
\begin{bmatrix} \widehat{\boldsymbol{\theta}}_{\mathbf{A}}^{[i+1]} \\ \widehat{\overline{\boldsymbol{\lambda}}}^{[i+1]} \end{bmatrix} &= \begin{bmatrix} \widehat{\boldsymbol{\theta}}_{\mathbf{A}}^{[i]} \\ \widehat{\overline{\boldsymbol{\lambda}}}^{[i]} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_i^T \widetilde{\mathbf{W}} \mathbf{H}_i & \mathbf{H}_i^T \widetilde{\mathbf{W}} \overline{\mathbf{G}}_i \\ \overline{\mathbf{G}}_i^T \widetilde{\mathbf{W}} \mathbf{H}_i & \overline{\mathbf{G}}_i^T \widetilde{\mathbf{W}} \overline{\mathbf{G}}_i \end{bmatrix}^{-1} \\
&\cdot \begin{bmatrix} \mathbf{H}_i^T \widetilde{\mathbf{W}} \\ \overline{\mathbf{G}}_i^T \widetilde{\mathbf{W}} \end{bmatrix} \left(\widetilde{\mathbf{y}} - \widetilde{\mathbf{G}}_o(\widehat{\mathbf{A}}^{[i]})\widehat{\boldsymbol{\lambda}}^{[i]}\right) \qquad i = 0, 1, \ldots \quad (22)
\end{aligned}$$

where

$$\mathbf{H}_i = \left.\frac{\partial \widetilde{\mathbf{G}}_o(\mathbf{A})\boldsymbol{\lambda}}{\partial \boldsymbol{\theta}_{\mathbf{A}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[i]}} \quad \text{and} \quad \overline{\mathbf{G}}_i = \left.\frac{\partial \widetilde{\mathbf{G}}_o(\mathbf{A})\boldsymbol{\lambda}}{\partial \overline{\boldsymbol{\lambda}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[i]}}.$$

Further, exploiting the block diagonality of $\widetilde{\mathbf{W}}$ (13), the right term of (22) could be rewritten as sums of smaller matrices as follows:

$$\begin{aligned}
&\left\{\sum_{\ell=1}^{d}\sum_{k=\ell}^{d} \begin{bmatrix} \{\mathbf{H}_i\}_{k\ell}^T \mathbf{W}_{k\ell} \{\mathbf{H}_i\}_{k\ell} & \{\mathbf{H}_i\}_{k\ell}^T \mathbf{W}_{k\ell} \{\overline{\mathbf{G}}_i\}_{k\ell} \\ \{\overline{\mathbf{G}}_i\}_{k\ell}^T \mathbf{W}_{k\ell} \{\mathbf{H}_i\}_{k\ell} & \{\overline{\mathbf{G}}_i\}_{k\ell}^T \mathbf{W}_{k\ell} \{\overline{\mathbf{G}}_i\}_{k\ell} \end{bmatrix}\right\}^{-1} \\
&\cdot \left\{\sum_{\ell=1}^{d}\sum_{k=\ell}^{d} \begin{bmatrix} \{\mathbf{H}_i\}_{k\ell}^T \mathbf{W}_{k\ell} \left(\widetilde{\mathbf{y}}_{k\ell} - \mathbf{G}_{k\ell}(\widehat{\mathbf{A}}^{[i]})\widehat{\boldsymbol{\lambda}}^{[i]}\right) \\ \{\overline{\mathbf{G}}_i\}_{k\ell}^T \mathbf{W}_{k\ell} \left(\widetilde{\mathbf{y}}_{k\ell} - \mathbf{G}_{k\ell}(\widehat{\mathbf{A}}^{[i]})\widehat{\boldsymbol{\lambda}}^{[i]}\right) \end{bmatrix}\right\} \quad (23)
\end{aligned}$$

where $\widetilde{\mathbf{y}}_{k\ell}$, $\mathbf{W}_{k\ell}$ and $\mathbf{G}_{k\ell}(\mathbf{A})$ were defined in (14), (15), and (16), respectively, and

$$\{\mathbf{H}_i\}_{k\ell} = \left.\frac{\partial \mathbf{G}_{k\ell}(\mathbf{A})\boldsymbol{\lambda}}{\partial \boldsymbol{\theta}_{\mathbf{A}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[i]}} \qquad M \times d^2 \qquad (24)$$

$$\{\overline{\mathbf{G}}_i\}_{k\ell} = \left.\frac{\partial \mathbf{G}_{k\ell}(\mathbf{A})\boldsymbol{\lambda}}{\partial \overline{\boldsymbol{\lambda}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[i]}} \qquad M \times (M-1)d$$

$$= \mathbf{G}_{k\ell}(\mathbf{A}^{[i]})[\mathbf{0}\vdots\mathbf{I}_{d(M-1)}]^T \qquad (25)$$

The initial conditions could be set as follows: $\widehat{\boldsymbol{\theta}}_{\mathbf{A}}^{[0]} = \mathrm{vec}\{\mathbf{I}_d\}$ and $\widehat{\boldsymbol{\lambda}}_{\tau}^{[0]} = \mathrm{diag}(\widehat{\mathbf{R}}_{\mathbf{s}}[\tau])$ for $\tau = 1, \ldots, M-1$, where "diag" denotes the diagonal elements of $\widehat{\mathbf{R}}_{\mathbf{s}}[\tau]$. Usually, 3 iterations are sufficient to achieve convergence of the algorithm.

Note that the size of the matrix to be inverted in (23) is $d(d+M-1) \times d(d+M-1)$. Computation of the matrix and its inversion involves $O(d^4M^2)$ and $O(d^6 + d^3M^3)$ operations, respectively. Hence we can conclude that the total cost of the minimization of the weighted criterion is $O(d^6 + d^3M^3)$ operations.

3. COMPLEXITY ISSUES

For evaluating the computational complexity of the algorithm, we can divide it into four parts.

1. Computation of the lagged covariance matrices requires d^2MN operations.
2. Computation of SOBI - requires $O(d^4M)$ operations.
3. Computation of diagonal blocks $\mathbf{W}_{k\ell}$, $k, \ell = 1, \ldots, d$, $k \neq \ell$ of the weight matrix $\widetilde{\mathbf{W}}$. Each (k, ℓ) pair requires to evaluate $2M$ values of the function ϕ, each requiring $O(M^2)$ operations, all together it is $O(d^2M^3)$ operations to form $\frac{d(d+1)}{2}$ $M \times M$ blocks. Inversion of each block requires $O(M^3)$ operations, which does not alter the complexity $O(d^2M^3)$ operations.
4. Minimization of the criterion (10) is done recursively, each iteration requiring $O(d^6 + d^3M^3)$ operations.

4. CRAMÉR-RAO LOWER BOUND

As mentioned in the Section 2, the performance of BSS algorithms can be measured by means of the ISR matrix. The $(k, \ell)-$th element of this matrix $\mathrm{ISR}_{k\ell}$ characterizes the mean square residual presence of the $k-$th original source in the $\ell-$th estimated source. For both theoretical and practical reason it is interesting to compare the ISR obtained in simulations with the best achievable lower bound, in particular the CRB. For Gaussian parametric AR and ARMA sources, the bound was derived in general form in [10]. We cannot include more details here for lack of space, but it can be shown that in the case of AR sources, the CRB on ISR can be written in the form

$$\mathrm{ISR}_{k\ell} \geq \mathrm{CRLB}_{k\ell} = \frac{1}{N} \frac{\phi_{k\ell}}{1 - \phi_{k\ell}\phi_{\ell k}} \frac{\frac{1}{\sigma_\ell^2}\mathbf{R}_\ell[0]}{\frac{1}{\sigma_k^2}\mathbf{R}_k[0]} \qquad (26)$$

where $\{\mathbf{R}_k[t]\}_{t=0}^{M-1}$ is the covariance sequence of the $k-$th source, σ_k^2 is the variance of the innovation sequence of the source,

$$\phi_{k\ell} = \frac{1}{\sigma_k^2} \sum_{i,j=0}^{M-1} a_{i\ell} a_{j\ell} \mathbf{R}_k[i-j] = \frac{1}{2\pi j} \oint \frac{A_\ell(z) A_\ell^*(z) z^{-1}}{A_k(z) A_k^*(z)} dz$$

$\{a_{i\ell}\}_{i=0}^{M-1}$ are the AR coefficients of the $\ell-$th source with $a_{0\ell} = 1$ $\forall \ell$, and $A_\ell(z) = \sum_{i=0}^{M} a_{i\ell} z^{-i}$, $\ell = 1, \ldots, d$.

Since WASOBI is asymptotically equivalent to Maximum Likelihood ([10]), it asymptotically attains the CRB, which can be estimated in turn from the data and serve to predict the performance of WASOBI.

5. SIMULATIONS

Performance of the new implementation of the WASOBI was tested on simulated data of the length $N = 5000$ with $d = 20$ components. The first 10 component were constructed so that the $k-$th component was Gaussian AR process of the $k-$th order, with AR parameters $(1, 0, \ldots, 0, \rho)$, where ρ was a free parameter in interval $(0, 1)$, and $k = 1, \ldots, 10$. The remaining 10 components were defined similarly with the difference that the AR parameters were $(1, 0, \ldots, 0, -\rho)$. The parametric choice of the sources has the advantage that by controlling a single parameter (ρ), it allows us to simulate sources that are difficult to separate (for ρ close to zero), and sources that are relatively easy to separate (for ρ close to one). In the former case, the power spectra of the sources are similar to each other, whereas in the latter case the spectra are very different from each other.

All these sources can be characterized by the same "spectral dynamic range" $\eta \triangleq 20 \log_{10}\left(\frac{1+|\rho|}{1-|\rho|}\right)$, which is the ratio (in [dB]) between the peak spectral density of the sources and their minimal spectral density. Note that $\eta = -\infty$ for white processes ($\rho = 0$) and extends to infinity as ρ approaches ± 1.

Performance of the WASOBI is compared with that of SOBI and with the corresponding Cramér Rao bound in terms of the interference-to-signal ratio (ISR) for various η in Figure 1. The empirical ISR is based on 10 independent trials for each η. We can see that (1) WASOBI is nearly statistically efficient (attains the CRB) in the whole range of η between 0 dB and 40 dB 2) WASOBI outperforms SOBI significantly in terms of ISR, namely for moderate and high η.

5.1 Sensitivity to additive noise

To evaluate its robustness, WASOBI was also tested in the presence of white Gaussian additive noise (in the same scenario). In accordance with the recommendation in [11], we study the achieved signal-to-interference plus noise, SINR, and compare it with the SINR obtained by the hypothetical separator computed for known mixing matrix and variance of the noise. Results are shown in Figures 2 and 3. They demonstrate good performance of WASOBI in the noisy environment, but show, that if the additive noise exceeds a certain level, the difference in performance between SOBI and WASOBI disappears. In addition, we have not observed any improvement of performance of SOBI or SOBI-RO, if the number of the

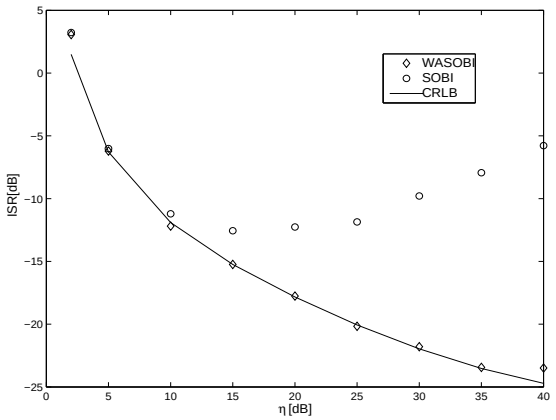

Figure 1: Average interference-to-signal ratio achieved in simulation by WASOBI and SOBI as a function of the spectral dynamic range η of the 20 sources.

involved matrices was increased above $M = 11$. Performance of SOBI-RO did not improve even at M as high as 100.

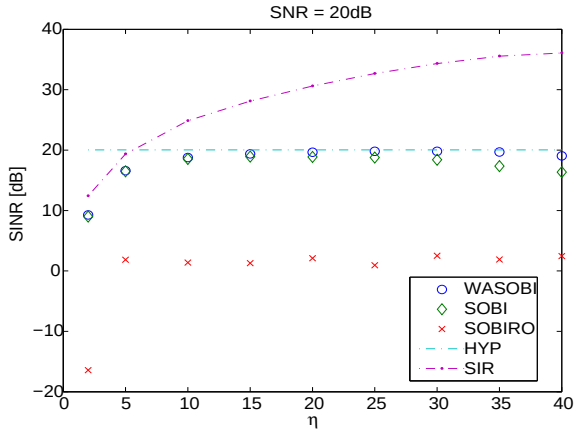

Figure 2: Average SINR achieved in simulation by WASOBI and SOBI where the mixture is contaminated by additive white Gaussian noise at SNR=20dB as a function of the spectral dynamic range η of the sources.

6. CONCLUSIONS

A novel implementation of the weighted SOBI (WASOBI) algorithm was proposed. At the present time, it allows a separation of about 20 sources in Matlab computation time in order of minutes. The algorithm is asymptotically efficient, which means that it attains the performance limit given by the Cramér-Rao lower bound if the data length is large enough, provided that the initial separation - here done by SOBI - is sufficiently good. Matlab code for the proposed implementation of WASOBI is available at `http://si.utia.cas.cz/Tichavsky.html`.

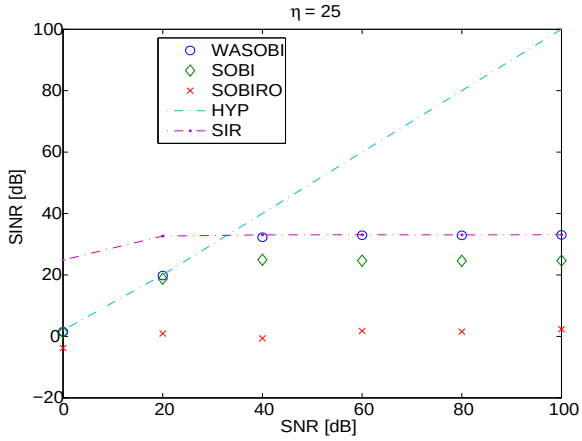

Figure 3: Average SINR achieved in simulation by WASOBI and SOBI for $\eta = 25dB$ in a noisy scenario as a function of the input SNR.

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, February 1997.

[2] C.A. Joyce , I.F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation", *Psychophysiology* **41** (2): 313-325, March 2004.

[3] C.J.James and C.W. Hesse, "Independent component analysis for biomedical signals", *Physiological Measurements*, vol. 26, R15-R39, 2005.

[4] L. Shoker, S. Sanei, J. Chambers, "Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm", *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 721-724, October 2005.

[5] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. West Sussex, UK: John Wiley & Sons, 2003.

[6] S. A. Cruces-Alvarez and A. Cichocki, "Combining blind source extraction with joint approximate diagonalization: thin algorithms for ICA," in *Proc. ICA 2003*, Nara, Japan, April 1-4, 2003. pp. 173-178.

[7] A. Belouchrani and A. Cichocki, "Robust whitening procedure in blind source separation context," *Electronics Letters*, vol. 36, no. 24, pp. 2050-2053, 2000.

[8] A. Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *IEEE Signal Processing Letters*, vol. 7, pp. 197-200, July 2000.

[9] E. Doron and A. Yeredor, "Asymptotically Optimal Blind Separation of Parametric Gaussian sources", *Proc. of ICA 2004*, C.G. Puntonet and A. Prieto (Eds.), pp. 390-397, 2004.

[10] E. Doron, "Asymptotically Optimal Blind Separation of Parametric Gaussian Sources", M.Sc. Thesis, Tel-Aviv University, 2003.

[11] Z. Koldovský, P. Tichavský "Methods of fair comparison of performance of linear ICA techniques in presence of additive noise", *Proc. ICASSP 2006*, Toulouse, France, May 14-19, 2006.

[12] B. Porat, *Digital Processing of Random Signals: Theory and Methods*, Prentice Hall, 1994.

[13] H. W. Sorenson, *Parameter estimation*, New York:Dekker,1980.