USING FARTHER CORRELATIONS TO FURTHER IMPROVE THE OPTIMALLY-WEIGHTED SOBI ALGORITHM

Arie Yeredor and *Eran Doron*
Department of EE-Systems, Tel-Aviv University
P.O.Box 39040, Tel-Aviv 69978, Israel
Tel: +972 3 6405314; fax: +972 3 6407095
e-mail: arie@eng.tau.ac.il, erandoron@hotmail.com

ABSTRACT

The Weights-Adjusted Second-Order Blind Identification (WASOBI) algorithm was recently proposed (Yeredor, 2000) as an optimized version of the SOBI Algorithm (Belouchrani et al., 1997) for blind separation of static mixtures of Gaussian Moving Average (MA) sources. The optimization consists of transforming the approximate joint diagonalization in SOBI into a properly weighted Least-Squares problem, with the asymptotically optimal weights specified in terms of the estimated correlations. However, only correlations up to the lag of the maximal MA order were used. Somewhat counter-intuitively, it turns out that estimated correlation matrices beyond this lag are also useful, although the respective true correlations are known to be zero and have no direct dependence on the mixing matrix. Nevertheless, when properly incorporated into the weighted least-squares problem, these estimated matrices can significantly improve performance, since they bear information on the estimation errors of the shorter-lags matrices. In this paper we show how to modify the WASOBI algorithm accordingly, and demonstrate the improvement via analysis and simulation results.

1 INTRODUCTION

Blind Source Separation (BSS) involves estimation of the mixing matrix $\boldsymbol{A}$ in the following mixture model

$$\boldsymbol{x}[t] = \boldsymbol{A}\boldsymbol{s}[t] \quad t = 1, 2, \ldots T, \tag{1}$$

where $\boldsymbol{s}[t] = \left[s_1[t]\; s_2[t]\; \cdots\; s_N[t]\right]^T$ are N unknown, statistically independent source signals, $\boldsymbol{x}[t] = \left[x_1[t]\; x_2[t]\; \cdots\; x_M[t]\right]^T$ are the M observations and $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ is the unknown mixing matrix. The term 'blind' ascribes lack of any additional information regarding the signals or $\boldsymbol{A}$.

In [1], Belouchrani et al. proposed the "Second-Order Blind Identification" (SOBI) algorithm for stationary source signals with distinct spectra, based on the joint diagonalization property of the observations' correlation matrices $\boldsymbol{R}_x[\tau] \stackrel{\triangle}{=} E\left[\boldsymbol{x}[t+\tau]\boldsymbol{x}^H[t]\right] = \boldsymbol{A}\boldsymbol{R}_s[\tau]\boldsymbol{A}^H$, where $\boldsymbol{R}_s[\tau] \stackrel{\triangle}{=} E\left[\boldsymbol{s}[t+\tau]\boldsymbol{s}^H[t]\right]$ are the source signals' (unknown) diagonal correlation matrices.

Following estimates of the observations' correlation matrices at different lags, SOBI seeks the approximate joint diagonalization of these matrices in two phases: The first phase is a "whitening" phase, in which the observations are spatially whitened, such that the empirical (estimated) correlation matrix at lag zero equals the identity matrix. Subsequently, a unitary matrix is found, using, e.g., successive Jacobi rotations ([1], [3]), such that the set of whitened empirical correlation matrices at nonzero lags are most closely diagonalized. The estimated mixing matrix $\hat{\boldsymbol{A}}$ is then given by the product of the unitary matrix and the inverse of the whitening matrix.

The procedure described above can be regarded as an attempt to attain a Least-Squares (LS) fit of the estimated correlation matrices in terms of the unknown parameters (the mixing matrix and the sources' correlations). However, it also implies some special weighting of the LS criterion, which is generally far from optimal. It was therefore proposed in [5], [6] to transform the joint diagonalization into a properly weighted LS (WLS) problem. In the case of source signals which are Gaussian Moving-Average (MA) processes of known maximal order (denoted Q), the optimal weight matrix can be expressed in terms of the observations' correlation matrices up to lag Q. Consequently, since all of these matrices would usually be estimated as the set to be jointly diagonalized, these estimates can also serve for the secondary purpose of computing the optimal weight matrix. Since the correlation estimates are consistent, the asymptotic behavior of such a scheme would approach the use of the true (optimal) weight matrices. This algorithm was termed the "Weights-Adjusted SOBI" (WASOBI).

Although the number of estimated matrices (actually the number of lags) to be used was not restricted in [5], [6], both the analysis and simulations only addressed lags less or equal to Q. This was mostly based on the property, that for MA processes the true correlations at farther lags are all zeros, and hence do not contain any information on the mixing matrix. Nevertheless, it turns out that although the true (all zeros) correlation matrices at farther lags are "useless", their *estimated* counterparts, when estimated from the same data used for estimating the "useful" matrices, bear information on the estimation errors of these "useful" matrices. Therefore, exploiting the correlations between the estimation error of the "useless" matrices and the estimation errors of the "useful" matrices in the WLS framework can significantly improve the performance. A similar observation was recently proposed in [4] in the context of spectral estimation of MA processes.

However, straightforward extension of WASOBI, to merely use correlation estimates up to lag $Q' > Q$ would not fully exploit the knowledge that true correlations at farther lags are all zeros. This information has to be incorporated into the LS model, in order to prevent the algorithm from attempting to jointly diagonalize the extra matrices as well. In addition, the information for the weight matrix should only be extracted from the first Q estimated matrices, and not from all Q' estimated matrices.

In this paper we derive the extended WASOBI algorithm, which uses $Q' > Q$ matrices, but keeps the MA model order fixed at Q. The derivation also enables to obtain analytic expressions for the resulting asymptotic performance. We present analytic results, and compare to SOBI and WASOBI using simulation results as well.

We shall focus on the case $M = N = 2$ real-valued signals with a real-valued mixing matrix. Extension to the more general case is straightforward, but requires some more complicated, extended notations, which we choose to avoid in here, since they are irrelevant to the essence of the algorithm.

2 FORMULATION AS A WEIGHTED LS PROBLEM

We assume throughout, that the source signals are all MA processes, whose orders are known to be less or equal to Q. The estimated correlation matrices are denoted $\hat{\boldsymbol{R}}_x[k]$ for $k = 0, 1, ...Q'$, where $Q' > Q$:

$$\hat{\boldsymbol{R}}_x[k] = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}[t]\boldsymbol{x}^T[t+k] \quad k = 0, 1, ...Q'. \qquad (2)$$

Note that since all matrices are estimated from T samples, (2) assumes implicitly that the actual number of available samples is $T + Q'$. This is somewhat wasteful in the sense that not all available data points are used for the smaller lags. However, this wastefulness is negligible when T is large relative to Q', and this assumption simplifies the derivation of optimal weights in the next section.

Formulation of a LS model requires the description of the available (inaccurate) measurements in terms of the parameters of interest. We use the elements of the estimated correlation matrices as the set of raw measurements. We seek a 2×2 matrix $\boldsymbol{A}$ and $Q + 1$ diagonal matrices $\boldsymbol{\Lambda}_0, \boldsymbol{\Lambda}_1 \ldots \boldsymbol{\Lambda}_Q$ such that $\hat{\boldsymbol{R}}_x[k]$ are "best fitted" by $\boldsymbol{A}\boldsymbol{\Lambda}_k\boldsymbol{A}^T$ for $k = 0, 1, \ldots Q$ (only), whereas the other $Q' - Q$ matrices are fitted by zeros. Thus, there are four parameters of interest, denoted $\boldsymbol{a} \triangleq \mathrm{vec}\{\boldsymbol{A}\} = [\boldsymbol{A}^{(1,1)}\boldsymbol{A}^{(2,1)}\boldsymbol{A}^{(1,2)}\boldsymbol{A}^{(2,2)}]^T$, and $2(Q+1)$ nuisance parameters, which are the $Q + 1$ 2×1 vectors $\boldsymbol{\lambda}_k \triangleq \mathrm{diag}\{\boldsymbol{\Lambda}_k\}$ $k = 0, 1, \ldots Q$. However, due to the inherent scaling ambiguity (which enables to commute scales between $\boldsymbol{A}$ and $\boldsymbol{\Lambda}_k$), we may arbitrarily fix e.g. $\boldsymbol{\Lambda}_0$, reducing the true number of nuisance parameters to $2Q$.

Note that the estimated $\hat{\boldsymbol{R}}_x[\tau_k]$ 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 $\hat{\boldsymbol{r}}_k \triangleq \mathrm{vec}\{\hat{\boldsymbol{R}}_x[\tau_k]\}$ and

$$\boldsymbol{y}_k \triangleq \boldsymbol{C}\hat{\boldsymbol{r}}_k \quad k = 0, 1, \ldots Q', \qquad (3)$$

where $\boldsymbol{C}$ is a constant transformation matrix,

$$\boldsymbol{C} \triangleq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad (4)$$

are the actual measurements of the LS model. The desired fit for each k can then be written as

$$\boldsymbol{y}_k \approx \boldsymbol{G}(\boldsymbol{a})\boldsymbol{\lambda}_k. \qquad (5)$$

where the matrix $\boldsymbol{G}(\boldsymbol{a})$ is given by

$$\boldsymbol{G}(\boldsymbol{a}) \triangleq \begin{bmatrix} a_1^2 & a_3^2 \\ a_1a_2 & a_3a_4 \\ a_2^2 & a_4^2 \end{bmatrix}. \qquad (6)$$

Concatenating the first $Q + 1$ $\boldsymbol{y}_k$-s into $\boldsymbol{y} \triangleq [\boldsymbol{y}_0^T\boldsymbol{y}_1^T \cdots \boldsymbol{y}_Q^T]^T$, and the other $Q' - Q$ $\boldsymbol{y}_k$-s into $\tilde{\boldsymbol{y}} \triangleq [\boldsymbol{y}_{Q+1}^T\boldsymbol{y}_{Q+2}^T \cdots \boldsymbol{y}_{Q'}^T]^T$, we get

$$\boldsymbol{y} \approx [\boldsymbol{I}_{Q+1} \otimes \boldsymbol{G}(\boldsymbol{a})]\boldsymbol{\lambda} \triangleq \tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda} \qquad (7)$$

$$\tilde{\boldsymbol{y}} \approx \boldsymbol{0} \qquad (8)$$

where $\boldsymbol{I}_{Q+1}$ denotes the $(Q+1)\times(Q+1)$ identity matrix, $\otimes$ denotes Kronecker's product, $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_0^T\boldsymbol{\lambda}_1^T \cdots \boldsymbol{\lambda}_Q^T]^T$ is the concatenation of $\boldsymbol{\lambda}_k$, and $\boldsymbol{0}$ denotes a $3(Q' - Q) \times 1$ all-zeros vector. We also define

$$\bar{\boldsymbol{\lambda}} = [\boldsymbol{\lambda}_1^T\boldsymbol{\lambda}_2^T \cdots \boldsymbol{\lambda}_Q^T]^T, \qquad (9)$$

the vector of free parameters in $\boldsymbol{\lambda}$.

Given any $3(Q' + 1) \times 3(Q' + 1)$ symmetric weight matrix $\boldsymbol{W}$, we may now define the WLS criterion as

$$C_{WLS}(\boldsymbol{a}, \boldsymbol{\lambda}) \triangleq \begin{bmatrix} \boldsymbol{y} - \tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda} \\ \tilde{\boldsymbol{y}} \end{bmatrix}^T \boldsymbol{W} \begin{bmatrix} \boldsymbol{y} - \tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda} \\ \tilde{\boldsymbol{y}} \end{bmatrix}. \qquad (10)$$

If we further partition $\boldsymbol{W}$ into

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_{11} & \boldsymbol{W}_{12} \\ \boldsymbol{W}_{12}^T & \boldsymbol{W}_{22} \end{bmatrix} \tag{11}$$

where $\boldsymbol{W}_{11}$ is $(Q+1) \times (Q+1)$, then C_{WLS} can be expressed as

$$\begin{aligned} C_{WLS}(\boldsymbol{a}, \boldsymbol{\lambda}) = [\boldsymbol{y} - \tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda}]^T \boldsymbol{W}_{11} [\boldsymbol{y} - \tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda}] \\ +2[\boldsymbol{y} - \tilde{\boldsymbol{G}}(\boldsymbol{a})\boldsymbol{\lambda}]^T \boldsymbol{W}_{12}\tilde{\boldsymbol{y}} + \tilde{\boldsymbol{y}}^T \boldsymbol{W}_{22}\tilde{\boldsymbol{y}}, \end{aligned} \tag{12}$$

to be minimized with respect to (w.r.t.) $\boldsymbol{a}$ and $\bar{\boldsymbol{\lambda}}$, with $\boldsymbol{\lambda}_1$ set arbitrarily. While linear (quadratic) in $\bar{\boldsymbol{\lambda}}$, this WLS criterion is nonlinear in $\boldsymbol{a}$. Several methods for minimizing C_{WLS} can be considered. For example, Gauss iterations (see e.g. [7]) can be used. However, To exploit the linear part (w.r.t. $\bar{\boldsymbol{\lambda}}$), the Gauss iterations may be restricted to the nonlinear minimization w.r.t. $\boldsymbol{a}$ with $\bar{\boldsymbol{\lambda}}$ fixed. Thus, C_{WLS} can be minimized by alternating between linear (closed-form) minimization w.r.t. $\bar{\boldsymbol{\lambda}}$ with $\boldsymbol{a}$ fixed, and vice-versa. Another appealing approach would be to interlace minimizations w.r.t. $\bar{\boldsymbol{\lambda}}$ with the Gauss iterations. The SOBI estimate may be used as an initial value for the iterations.

3 OPTIMAL WEIGHTING

The LS criterion presented above allows the use of any (arbitrary) weight matrix $\boldsymbol{W}$. Naturally, we would like to use the optimal weight matrix, which is well-known (e.g. [7]) to be the inverse of the measurements' covariance matrix. Thus we need the covariance matrix of the entire "measurements" vector $[\boldsymbol{y}^T \tilde{\boldsymbol{y}}^T]$, which we shall denote $\boldsymbol{\Phi}$. It is only through the use of this augmented covariance matrix, that the information in $\tilde{\boldsymbol{y}}$ can become useful in improving an estimate based merely on $\boldsymbol{y}$. Assuming Gaussian signals, we have from (2)

$$\begin{aligned} & E\left[\hat{\boldsymbol{R}}_x^{(i,j)}[k]\hat{\boldsymbol{R}}_x^{(m,n)}[l]\right] = \\ & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[x_i[t]x_j[t+k]x_m[s]x_n[s+l]\right] = \\ & \boldsymbol{R}_x^{(i,j)}[k]\boldsymbol{R}_x^{(m,n)}[l] + \\ & \frac{1}{T}\sum_{p=-(T-1)}^{T-1}(1-\frac{|p|}{T})\boldsymbol{R}_x^{(i,m)}[p]\boldsymbol{R}_x^{(j,n)}[p+l-k] + \\ & \frac{1}{T}\sum_{p=-(T-1)}^{T-1}(1-\frac{|p|}{T})\boldsymbol{R}_x^{(i,n)}[p+l]\boldsymbol{R}_x^{(j,m)}[p-k] \\ & \qquad k,l = 0,1,\dots Q' \end{aligned} \tag{13}$$

which implies that the covariance of $\hat{\boldsymbol{R}}_x^{(i,j)}[k]$ and $\hat{\boldsymbol{R}}_x^{(m,n)}[l]$ is given by sum of the last two terms. Now, since the source signals are MA of orders $\leq Q$, we have $\boldsymbol{R}_x[p] = \boldsymbol{0}\ \forall |p| > Q$; since the lags k,l are all non-negative, there is at least one zero factor in each expression for $|p| > Q$, so the summation over p can be reduced to $-Q$ to Q. Consequently, estimating the correlation matrices up to lag Q is sufficient for consistently estimating $\boldsymbol{\Phi}$. The use of the other estimated matrices, for lags between Q and Q' would only serve for augmenting the WLS criterion, but not for determining the weight matrix.

With slight manipulations (13) can be reformulated in matrix form as

$$\begin{aligned} & Cov[\hat{\boldsymbol{r}}_k, \hat{\boldsymbol{r}}_l] = \\ & \frac{1}{T}\sum_{p=-Q}^{Q}(1-\frac{|p|}{T})\boldsymbol{R}_x[p+\tau_l-\tau_k] \otimes \boldsymbol{R}_x[p] + \\ & \frac{1}{T}\sum_{p=-Q}^{Q}(1-\frac{|p|}{T})(\boldsymbol{R}_x[p-\tau_k] \otimes \boldsymbol{R}_x[p-\tau_l])\boldsymbol{P} \end{aligned} \tag{14}$$

where $\boldsymbol{P}$ is a permutation matrix that swaps the second and third columns of the matrix to its left. Recalling the linear transformation (5) 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

$$\boldsymbol{\Phi}_{k,l} \triangleq Cov[\boldsymbol{y}_k, \boldsymbol{y}_l] = \boldsymbol{C} Cov[\hat{\boldsymbol{r}}_k, \hat{\boldsymbol{r}}_l]\boldsymbol{C}^T. \tag{15}$$

The optimal weight matrix is then given by $\boldsymbol{W}_{opt} = \boldsymbol{\Phi}^{-1}$. In practice, estimated correlations would replace true correlations in (14), providing a consistent estimate of $\boldsymbol{W}_{opt}$. Thus the resulting weights are asymptotically optimal. Under non-asymptotic condition, however, the estimated covariance matrix $\boldsymbol{\Phi}$ may be ill-conditioned, or even sign-indefinite, having very small (negative or positive) eigenvalues. In order to avoid numerical problems, it is recommended under such conditions to artificially improve the conditioning of $\boldsymbol{\Phi}$ before inversion by adding a small constant to its diagonal.

4 SIMULATIONS RESULTS

Fig. 1 presents some simulations results in terms of the mean Interference to Signal Ratio (ISR) for SOBI, WASOBI, and Extended WASOBI vs. the observation length T. The source signals used were MA(4) and MA(2) processes: $s_1(t)$ is an MA(4) process with zeros at $1.2e^{\pm j\frac{\pi}{2}}$, $0.9e^{\pm j\frac{\pi}{3}}$ (and their reciprocals); $s_2(t)$ is an MA(2) process with zeros at $-0.75e^{\pm j\frac{\pi}{2}}$ (and their reciprocals). We thus have $Q = 4$. The extended WASOBI algorithm was run with $Q' = 6$. All algorithms used the same data. Each simulation point represents an average of 400 trials. The mixing matrix used was $\boldsymbol{A} = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$.

To put the simulations results in context, we also present (in solid lines, superimposed on simulations results) the theoretically predicted performance: Since the "measurements" $\boldsymbol{y}$, $\tilde{\boldsymbol{y}}$ are unbiased (their expected value are the true correlation values), the estimated parameters are also unbiased, under a small-errors assumption (regardless of the weighting used). Using the derivative

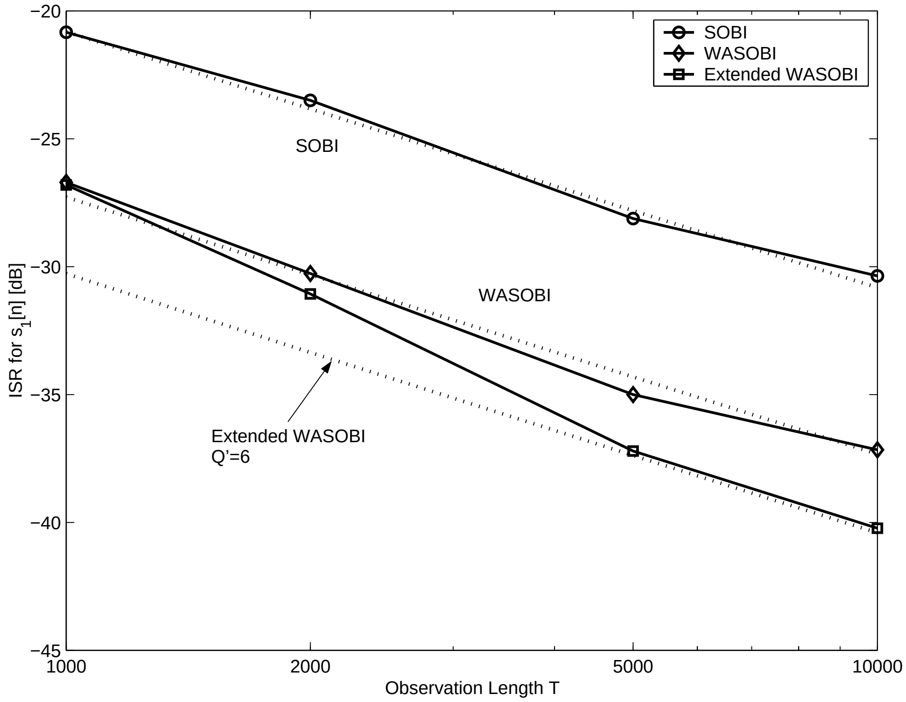

Figure 1: Simulations results (and theoretically predicted results) for SOBI, WASOBI and Extended WASOBI (with $Q' = 6$) in terms of the ISR for $s_1[n]$, vs. the observation length T. Source signals are MA(4) and MA(3) Gaussian processes. Both algorithms used the same data. Each simulation point represents an average of 400 trials.

of the LS criterion (10) with respect to all the parameters, as well as the measurements' covariance $\mathbf{\Phi}$, standard tools can be used (e.g. [7]) to obtain the (approximate) error covariance in estimating $\boldsymbol{a}$, the elements of $\boldsymbol{A}$. This covariance can in turn be translated to the mean ISR obtained when the estimated $\boldsymbol{A}$ is used for reconstruction of the source signals. See [5] for an explicit derivation. The resulting expressions are general, and can be used with any weight matrix $\boldsymbol{W}$. The analytic results for WASOBI and Extended WASOBI were calculated using the true optimal weight matrix, whereas for the simulations results the estimated optimal weight matrix was used. It is seen, as expected, that asymptotically the simulation values coincide with the predicted values. For SOBI, a weight matrix attributing high weight for the zero-lag correlations was used to predict the performance. The simulations results are seen to asymptotically approach their theoretical values.

5 CONCLUSION

We have shown that when the optimally weighted WASOBI algorithm is used, its performance can be further improved by incorporating estimated correlation matrices from lags beyond the MA order. The fact that the true correlations at these lags all zero out is used for proper parameterization the LS model and for estimation of the (asymptotically) optimal weight matrix. Note that these additional matrices can only be exploited with the WASOBI algorithm through the use of proper weighting. They remain practically useless for the basic SOBI algorithm.

6 *

References

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

[2] J.-F. Cardoso, "On the Performance of Orthogonal Source Separation Algorithms" *Proceedings of EUSIPCO'94*, pp. 776–779, 1994.

[3] J.-F. Cardoso, "Jacobi angles for simultaneous diagonalization" *SIAM Journal on Matrix Analysis and Applications*, vol. 17 no. 1, pp. 161–164, 1996.

[4] P. Stoica, T. McKelvey and J. Mari, "MA Estimation in Polynomial Time" *IEEE Transactions on Signal Processing* vol. 48 no. 7, pp. 1999–2012, 2000.

[5] A. Yeredor, "Blind Separation of Gaussian Sources via Second-Order Statistics with Asymptotically Optimal Weighting", *IEEE Signal Processing Letters*, vol. 7 no. 7, pp. 197–200, 2000.

[6] A. Yeredor, "Optimization of a Second-Order Statistics Blind Separation Algorithm for Gaussian Signals", *Proceedings of EUSIPCO 2000.*

[7] H. W. Sorenson, *Parameter Estimation* Marcel-Dekker, 1980.