

OPTIMIZATION OF A SECOND-ORDER STATISTICS BLIND SEPARATION ALGORITHM FOR GAUSSIAN SIGNALS

Arie Yeredor

Department of EE-Systems, Tel-Aviv University

P.O.Box 39040, Tel-Aviv 69978, Israel

Tel: +972 3 6408292; fax: +972 3 6407095

e-mail: arie@eng.tau.ac.il

ABSTRACT

The Second-Order Blind Identification (SOBI) algorithm (Belouchrani et al., 1997) is aimed at blind separation of static mixtures of stationary source signals with distinct spectra. It uses approximate joint diagonalization of empirical correlation matrices to estimate the mixing matrix. We show that SOBI's performance can be improved by transforming the joint diagonalization into a properly weighted nonlinear Least Squares problem. In the case of Gaussian sources, the optimal weights can be estimated consistently from the empirical correlation matrices. We demonstrate the substantial improvement by analysis and simulations.

1 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

$$\mathbf{x}[t] = \mathbf{A}\mathbf{s}[t] \quad t = 1, 2, \dots, T, \quad (1)$$

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

In [1], Belouchrani et al. proposed the "Second-Order Blind Identification" (SOBI) algorithm for stationary signals with distinct spectra. SOBI is based on the joint diagonalization property of the observations' correlation matrices $\mathbf{R}_x[\tau] \triangleq E[\mathbf{x}[t+\tau]\mathbf{x}^H[t]]$. Specifically, these matrices satisfy

$$\mathbf{R}_x[\tau] = \mathbf{A}\mathbf{R}_s[\tau]\mathbf{A}^H \quad \forall \tau \quad (2)$$

where $\mathbf{R}_s[\tau] \triangleq E[\mathbf{s}[t+\tau]\mathbf{s}^H[t]]$ are the source signals' (unknown) diagonal correlation matrices. Thus, \mathbf{A} is a joint diagonalizer of any set of K matrices, $\{\mathbf{R}_x[\tau_1], \mathbf{R}_x[\tau_2], \dots, \mathbf{R}_x[\tau_K]\}$. It can be shown that if

all the source signals have distinct 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 \mathbf{A} as the joint diagonalizer of a set of estimated correlation matrices $\{\hat{\mathbf{R}}_x[\tau_1], \hat{\mathbf{R}}_x[\tau_2], \dots, \hat{\mathbf{R}}_x[\tau_K]\}$. 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 \mathbf{A} by resorting to *approximate* joint diagonalization, attained in [1] in two phases:

In the first phase a "whitening" matrix $\hat{\mathbf{W}}$ is found, such that $\hat{\mathbf{W}}\hat{\mathbf{R}}_x[0]\hat{\mathbf{W}}^H = \mathbf{I}$ (identity matrix). All the other matrices are then similarly transformed,

$$\tilde{\mathbf{R}}[\tau_k] = \hat{\mathbf{W}}\hat{\mathbf{R}}_x[\tau_k]\hat{\mathbf{W}}^H \quad k = 1, 2, \dots, K. \quad (3)$$

In the second phase the *unitary* approximate joint diagonalizer $\hat{\mathbf{U}}$ of the transformed set is found, using successive Jacobi rotations, which iteratively minimize the off-diagonal entries of the transformed matrices ([1, 3]).

The desired estimate is then given by $\hat{\mathbf{A}} = \hat{\mathbf{W}}^\# \hat{\mathbf{U}}$ (where $\hat{\mathbf{W}}^\#$ denotes the pseudo-inverse of $\hat{\mathbf{W}}$).

It can be easily observed (see also [6]) that the second phase optimizes a Least-Squares (LS) fit of the $\tilde{\mathbf{R}}[\tau_k]$ -s with respect to $\hat{\mathbf{U}}$. However, this LS criterion is not optimized with respect to $\hat{\mathbf{A}}$, since the non-unitary part $\hat{\mathbf{W}}$ is chosen to attain *exact* diagonalization of $\hat{\mathbf{R}}_x[0]$, possibly at the expense of poor diagonalization of the other matrices. It has been noted in [2] that such "hard-whitening" operation bounds the attainable performance, since errors incurred by the estimated $\hat{\mathbf{R}}_x[0]$ and inflicted upon $\hat{\mathbf{W}}$ can no longer be compensated for by the other matrices, which only affect $\hat{\mathbf{U}}$.

Furthermore, while the "hard-whitening" approach imposes severely unbalanced weighting, it may be desired to not only balance the weights (e.g., use unweighted LS), but rather to seek the *optimal* weighting. In fact, since the errors in estimating the correlation values are themselves strongly correlated, it can be

expected that substantial improvement in performance may be attained by using an optimally weighted LS (WLS) criterion, which can account for these correlations by the use of a non-diagonal weight matrix. The weighting approach has also been proposed in [4], but not pursued further in there.

In this paper we address these two shortcomings of the SOBI algorithm. First, we reformulate the approximate diagonalization problem as a non-linear WLS problem. In that framework, several iterative algorithm can be considered for minimization with respect to an arbitrary (not necessarily unitary) matrix $\hat{\mathbf{A}}$. We then proceed to find the optimal weight matrix. In order to evaluate the correlation between the estimated correlation values, we assume that the source signals are Gaussian with finite-length correlations (i.e. Moving Average (MA) Processes). This assumption enables to express the optimal weight matrix in terms of the estimated correlation matrices, since all the required fourth-order moments can be expressed in terms of the available (estimated) second-order moments.

In conclusion we demonstrate (with 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. Our algorithm is given the acronym WASOBI (Weights-Adjusted SOBI).

2 FORMULATION AS A WEIGHTED LS PROBLEM

The formulation of a LS model requires the description of available (inaccurate) measurements in terms of the parameters of interest. We use the elements of the estimated correlation matrices at various (non-negative) lags as the set of raw measurements. For reasons that will become clear in the next section, we assume that all the estimated correlation matrices $\hat{\mathbf{R}}_x[\tau_k]$ are calculated using the same number T of data points. Specifically,

$$\hat{\mathbf{R}}_x[\tau_k] = \frac{1}{T} \sum_{t=1}^T \mathbf{x}[t] \mathbf{x}^T[t + \tau_k]. \quad (4)$$

Note that this assumes (implicitly) that $T + \tau_k$ samples are available for the estimation of $\mathbf{R}_x[\tau_k]$, which means that the actual number of available samples is $T + \tau_{max}$ where τ_{max} is the farthest lag used. This is somewhat wasteful in the sense that not all available data points are used for smaller lags. However, this wastefulness is negligible when T is large relative to τ_{max} , and this assumption simplifies the derivation of optimal weights in the next section.

We seek a 2×2 matrix \mathbf{A} and K diagonal matrices $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2 \dots \mathbf{\Lambda}_K$ such that $\hat{\mathbf{R}}_x[\tau_k]$ are "best fitted" by $\mathbf{A} \mathbf{\Lambda}_k \mathbf{A}^T$ for $k = 1, 2, \dots K$. Thus, there are

four parameters of interest, denoted $\mathbf{a} \triangleq \text{vec}\{\mathbf{A}\} = [\mathbf{A}^{(1,1)} \mathbf{A}^{(2,1)} \mathbf{A}^{(1,2)} \mathbf{A}^{(2,2)}]^T$, and $2K$ nuisance parameters, which are the $K \times 1$ vectors $\boldsymbol{\lambda}_k \triangleq \text{diag}\{\mathbf{\Lambda}_k\}$ $k = 1, 2, \dots K$. However, due to the inherent scaling ambiguity (which enables to commute scales between \mathbf{A} and $\mathbf{\Lambda}_k$), we may arbitrarily fix e.g. $\mathbf{\Lambda}_1$, reducing the true number of nuisance parameters to $2(K - 1)$.

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

$$\mathbf{y}_k \triangleq \mathbf{C} \hat{\mathbf{r}}_k \quad k = 1, 2 \dots K, \quad (5)$$

where \mathbf{C} is a constant transformation matrix,

$$\mathbf{C} \triangleq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (6)$$

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

$$\mathbf{y}_k \approx \mathbf{G}(\mathbf{a}) \boldsymbol{\lambda}_k. \quad (7)$$

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

$$\mathbf{G}(\mathbf{a}) \triangleq \begin{bmatrix} a_1^2 & a_3^2 \\ a_1 a_2 & a_3 a_4 \\ a_2^2 & a_4^2 \end{bmatrix}. \quad (8)$$

Concatenating all \mathbf{y}_k into $\mathbf{y} \triangleq [\mathbf{y}_1^T \mathbf{y}_2^T \dots \mathbf{y}_K^T]^T$, we get

$$\mathbf{y} \approx [\mathbf{I}_K \otimes \mathbf{G}(\mathbf{a})] \boldsymbol{\lambda} \triangleq \tilde{\mathbf{G}}(\mathbf{a}) \boldsymbol{\lambda} \quad (9)$$

where \mathbf{I}_K denotes the $K \times K$ identity matrix, \otimes denotes Kronecker's product, and $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1^T \boldsymbol{\lambda}_2^T \dots \boldsymbol{\lambda}_K^T]^T$ is the concatenation of $\boldsymbol{\lambda}_k$. We also define

$$\bar{\boldsymbol{\lambda}} = [\boldsymbol{\lambda}_2^T \boldsymbol{\lambda}_3^T \dots \boldsymbol{\lambda}_K^T]^T, \quad (10)$$

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

Given any $3K \times 3K$ symmetric weight matrix \mathbf{W} , we may now define the WLS criterion as

$$C_{WLS}(\mathbf{a}, \boldsymbol{\lambda}) \triangleq [\mathbf{y} - \tilde{\mathbf{G}}(\mathbf{a}) \boldsymbol{\lambda}]^T \mathbf{W} [\mathbf{y} - \tilde{\mathbf{G}}(\mathbf{a}) \boldsymbol{\lambda}] \quad (11)$$

to be minimized with respect to (w.r.t.) \mathbf{a} and $\bar{\boldsymbol{\lambda}}$, with $\boldsymbol{\lambda}_1$ set arbitrarily. While linear (quadratic) in $\bar{\boldsymbol{\lambda}}$, this WLS criterion is nonlinear in \mathbf{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. \mathbf{a} with $\bar{\boldsymbol{\lambda}}$ fixed. Thus, C_{WLS} can be minimized by alternating between linear (closed-form) minimization w.r.t.

$\bar{\lambda}$ with \mathbf{a} fixed, and vice-versa. Another appealing approach would be to interlace minimizations w.r.t. $\bar{\lambda}$ with the Gauss iterations. The SOBI estimate may be used as an initial value for the iterations.

The minimization of C_{WLS} would often be computationally more intensive than the SOBI minimization (this is obviously the case if the SOBI estimate is used as an initial guess). Note, however, that the computational load of the minimization depends only on K , and is independent of the number of observations T . Thus, asymptotically, the mean computational load per sample of the two methods is equal, since it is dominated by the calculation of the estimated correlation, rather than by the joint-diagonalization / LS-minimization.

3 OPTIMAL WEIGHTING

The LS criterion presented above allows the use of any (arbitrary) weight matrix \mathbf{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 \mathbf{y} , denoted Φ . Assuming Gaussian signals, we have from (4)

$$\begin{aligned} E \left[\hat{\mathbf{R}}_x^{(i,j)}[\tau_k] \hat{\mathbf{R}}_x^{(m,n)}[\tau_l] \right] = \\ \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T E \left[x_i[t] x_j[t + \tau_k] x_m[s] x_n[s + \tau_l] \right] = \\ \mathbf{R}_x^{(i,j)}[\tau_k] \mathbf{R}_x^{(m,n)}[\tau_l] + \\ \frac{1}{T} \sum_{p=-(T-1)}^{T-1} \left(1 - \frac{|p|}{T}\right) \mathbf{R}_x^{(i,m)}[p] \mathbf{R}_x^{(j,n)}[p + \tau_l - \tau_k] + \\ \frac{1}{T} \sum_{p=-(T-1)}^{T-1} \left(1 - \frac{|p|}{T}\right) \mathbf{R}_x^{(i,n)}[p + \tau_l] \mathbf{R}_x^{(j,m)}[p - \tau_k] \end{aligned} \quad (12)$$

which implies that the covariance of $\hat{\mathbf{R}}_x^{(i,j)}[\tau_k]$ and $\hat{\mathbf{R}}_x^{(m,n)}[\tau_l]$ is given by sum of the last two terms. We now further assume that the source signals are MA of orders $\leq Q$, whereas the selected lags are $\tau_k = k - 1$, $k = 1, 2, \dots, Q + 1$. The summation over p can then be reduced to $-Q$ 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 Φ .

With slight manipulations (12) can be reformulated in matrix form, such that

$$\begin{aligned} Cov[\hat{\mathbf{r}}_k, \hat{\mathbf{r}}_l] = \\ \frac{1}{T} \sum_{p=-Q}^Q \left(1 - \frac{|p|}{T}\right) \mathbf{R}_x[p + \tau_l - \tau_k] \otimes \mathbf{R}_x[p] + \\ \frac{1}{T} \sum_{p=-Q}^Q \left(1 - \frac{|p|}{T}\right) \mathbf{R}_x[p - \tau_k] \otimes \mathbf{R}_x[p - \tau_l] \mathbf{P} \end{aligned} \quad (13)$$

where \mathbf{P} is a permutation matrix that swaps the second and third columns of the matrix to its left. Recalling the linear transformation (7) from $\hat{\mathbf{r}}_k$ to \mathbf{y}_k we conclude that the (k, l) -th 3×3 block of Φ is given by

$$\Phi_{k,l} \triangleq Cov[\mathbf{y}_k, \mathbf{y}_l] = \mathbf{C} Cov[\hat{\mathbf{r}}_k, \hat{\mathbf{r}}_l] \mathbf{C}^T. \quad (14)$$

The optimal weight matrix is then given by $\mathbf{W}_{opt} = \Phi^{-1}$. In practice, estimated correlations would replace true correlations in (13), providing a consistent estimate of \mathbf{W}_{opt} . Thus the resulting weights are asymptotically optimal.

4 SIMULATIONS RESULTS

Fig. 1 presents some simulations results in terms of the mean Interference to Signal Ratio (ISR) for both SOBI and WASOBI, vs. the observation length T . The source signals used were MA(4) and MA(3) processes: $s_1(t)$ is an MA(4) process with zeros at $0.8e^{\pm j\frac{\pi}{4}}$, $0.7e^{\pm j\frac{\pi}{8}}$ (and their reciprocals); $s_2(t)$ is an MA(3) process with zeros at 0.7 , $0.3e^{\pm j\frac{\pi}{2}}$ (and their reciprocals). Both algorithms used the same data. Each simulation point represents an average of 1000 trials. The mixing matrix used was $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Interestingly, however, it turns out that performance (in terms of ISRs) does not depend on \mathbf{A} for *neither* SOBI *nor* WASOBI. The \mathbf{A} -invariance of SOBI agrees with [1]; for WASOBI it is more subtle to conclude from the derivation - note that such invariance is not attained with arbitrary \mathbf{W} !

To put the simulations results in context, we also present (in solid lines, superimposed on simulations results) the theoretically predicted performance: Since the "measurements" \mathbf{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 of the LS criterion (11) with respect to all the parameters, as well as the measurements' covariance Φ , standard tools can be used (e.g. [7]) to obtain the (approximate) error covariance in estimating \mathbf{a} , the elements of \mathbf{A} . This covariance can in turn be translated to the mean ISR obtained when the estimated \mathbf{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 \mathbf{W} .

For example, for the SOBI (unweighted) algorithm, \mathbf{W} would be set to $\mathbf{W}_{SOBI} = \text{diag}\{\alpha \mathbf{I}_3, \mathbf{I}_{3(K-1)}\}$ (in the matrices-to-matrix sense), where $\alpha \gg Q$ (we used $\alpha = 100$) is a large constant reflecting SOBI's obligatory whitening phase, which attributes infinite weight to warrant exact diagonalization of $\hat{\mathbf{R}}_x[0]$. For WASOBI we used the optimal weight, \mathbf{W}_{opt} . Note that while the *true* \mathbf{W}_{opt} was used for calculating the predicted performance, the WASOBI algorithm naturally used only the estimated \mathbf{W}_{opt} (based on the available data). It is seen that as expected, the predictions are approached as T

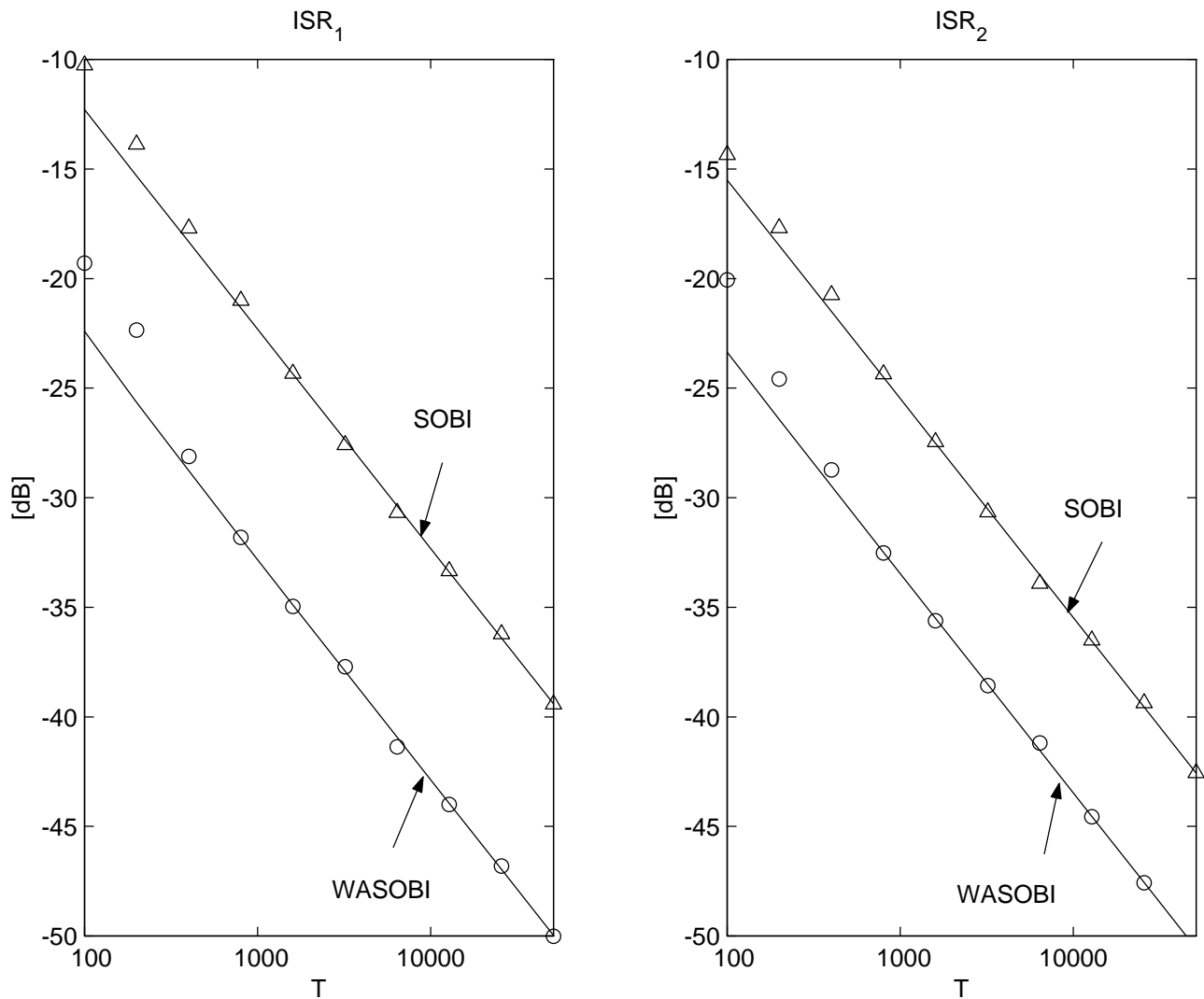


Figure 1: Simulations results (and theoretically predicted results) for SOBI and WASOBI, in terms of ISR, vs. the observation length T . Source signals are MA(4) and MA(3) Gaussian processes. Both algorithms used the same data. Each simulation point represents an average of 1000 trials.

increases, when the estimated \mathbf{W}_{opt} approaches the true \mathbf{W}_{opt} , and when the small-errors assumption prevails.

5 *

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] H. Sahlin, "Blind Signal Separation by Second Order Statistics" *PhD Thesis*, School of Electrical and Computer Engineering, Chalmers University of Technology, Sweden 1998.
- [5] A. Yeredor, "Blind Separation of Gaussian Sources via Second-Order Statistics with Asymptotically Optimal Weighting", accepted for publication in *IEEE Signal Processing Letters*, 2000.
- [6] M. Wax and J. Sheinvald, "A least-squares approach to joint diagonalization", *IEEE Signal Processing Letters*, vol. 4 no. 2, pp. 52–53, 1997.
- [7] H. W. Sorenson, *Parameter Estimation* Marcel-Dekker, 1980.