NEAR-OPTIMAL WEIGHTING IN CHARACTERISTIC-FUNCTION BASED ICA

Alon Slapak and *Arie Yeredor*

Dept. of Electrical Engineering – Systems, Tel-Aviv University
email: aslapak@post.tau.ac.il, arie@eng.tau.ac.il

ABSTRACT

In the context of Independent Component Analysis (ICA), we propose a near-optimal weighting scheme for the approximate joint diagonalization of empirical Hessians (second derivative matrices taken at selected "processing-points") of the observations' log-characteristic function. Our weighting scheme is based on the observation, that when the sources are nearly-separated, the covariance matrix of these empirical Hessians takes a convenient block-diagonal structure. We exploit this property to obtain reliable estimates of the blocks directly from the observed data, and use the recently proposed WEighted Diagonalization using Gauss itErations (WEDGE) to conveniently incorporate the weight matrices into the joint diagonalization estimation. Simulation results demonstrate the importance of proper weighting, especially for mitigating uncertainties in the selection of "processing points". As we show, the properly-weighted version can lead to a significant performance improvement, not only with respect to the unweighted version, but also with respect to a common benchmark like the popular JADE algorithm.

1. INTRODUCTION[1]

We consider the framework of Independent Component Analysis (ICA), where d statistically-independent (real-valued) sources are instantaneously mixed by an unknown constant mixing matrix $\mathbf{A}_0 \in \mathbb{R}^{d\times d}$, such that $\mathbf{x}[t]=\mathbf{A}_0\mathbf{s}[t]$, $1\le t\le T$ where $\mathbf{x}[t],\mathbf{s}[t]\in\mathbb{R}^{d\times 1}$ are the observations and sources vectors at time t. Given T observation vectors, the goal is to obtain an estimate $\hat{\mathbf{V}}$ of the demixing matrix, $\hat{\mathbf{V}}\approx\mathbf{A}_0^{-1}$, which in turn provides an estimate of the sources via $\hat{\mathbf{s}}[t]=\hat{\mathbf{V}}\mathbf{x}[t]$.

A popular approach to ICA is to apply approximate joint diagonalization (AJD) to a set of M matrix-form statistics (frequently termed "target-matrices"), having the appealing property of being strictly diagonal for random vectors with pairwise independent components. Some examples are the Joint Approximate Diagonalization of Eigenmatrices (JADE, [1]), whose target-matrices are derived from the observations' empirical fourth order cumulants; Second-Order Blind Identification (SOBI, [2]), whose target-matrices are the observations' empirical covariance matrices taken at different delay lags; And CHaracteristic-function-Enabled Source Separation (CHESS, [3]), whose target-matrices are empirical Hessians of the joint log-characteristic functions of the observations, taken at some selected "processing-points" (see more details in Section 2).

Since AJD can essentially be regarded as an attempted Least-Squares (LS) fit of the target-matrices to the joint diagonalization model, significant improvement of the resulting estimation accuracy can often be attained by applying proper weighting to the AJD process, thereby taking a Weighted LS (WLS) approach. The optimal weight matrix, assuming sufficiently small errors in the estimated target-matrices, is well-known to be given by the inverse of the joint covariance matrix of these errors. In general, however, this covariance matrix is unknown in a blind scenario, and must be obtained from the observed data.

Such an approach was first proposed in [4] for SOBI, where it was termed "Weights-Adjusted SOBI" (WASOBI), but the performance improvement in [4] was attained at the cost of a significant computational complexity. A simplified approach to the covariance estimation and to the subsequent weighted AJD was later proposed in [5,7], which offered a computationally efficient version of WASOBI. The simplifying idea is based on the observation that in a nearly-separated mixture, the required covariance matrix (and, hence, also the weight matrix) take a (nearly) block-diagonal form, and therefore admit more convenient manipulations. After an initial separation step, which is assumed to transform the demixed observation into a nearly-separated "mixture" of the sources, the demixing can be further refined by convenient estimation and application of the block-diagonal weights. A similar idea was used in [6] for applying optimal weighting in JADE.

Our goal in this paper is to use the same philosophy in applying asymptotically-optimal weighting to the CHESS algorithm. CHESS was shown in [3] to potentially attain significant performance improvement with respect to existing alternatives (such as JADE). However, the performance of CHESS generally depends on the specific choice of "processing-points" for the target-matrices. When an arbitrary selection of processing-points is used, the "bad" points (points at which the respective empirical Hessian has a large variance and little information content) can obscure the potential performance gain of the "good" points (at which the respective empirical Hessian has a small variance and rich information content). However, if proper (let alone optimal) weighting is used, the effect of the "bad" points would be out-weighted by that of the "good" points, and the desired performance gain would become evident. An initial weight-

[1] This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 1255/08).

 ISSN 2076-1465

ing scheme was presented in [8], but, like the initial WASOBI formulation [4], was extremely computationally demanding. Here we follow the computationally appealing "near-separation" philosophy of [5,7]. We rely on empirical estimation of the required covariance matrix (exploiting its block-diagonal structure) by partitioning the observation interval into blocks, followed by the efficient computation of the weighted AJD using the recently proposed Weighted Exhaustive Diagonalization using Gauss itErations (WEDGE, [5]).

The remainder of this paper is organized as follows: In section 2 we provide a brief overview of the principles of CHESS and of WEDGE. In section 3 we present the proposed weighting scheme in terms of the weights-estimation approach and the WEDGE-based reweighted AJD algorithm. Some experimental results are presented in Section 4 and conclusions are summarized in section 5.

2. PRELIMINARIES

For completeness of the exposition, we provide in this section a brief overview of CHESS [3] (for the basic principles of exploiting empirical off-origin Hessians of the log characteristic function), followed by a brief description of WEDGE [5] (for the weighted AJD stage).

2.1 CHESS – CHaracteristic-function-Enabled Source Separation

Let $\boldsymbol{\tau}$ denote a real-valued[2] arbitrarily-selected "processing-point" in $\mathbb{R}^{d\times 1}$. The Generalized Characteristic Function (GCF) of a random vector $\mathbf{s}\in\mathbb{R}^{d\times 1}$ at $\boldsymbol{\tau}$ is defined as $\varphi_{\mathbf{s}}(\boldsymbol{\tau})\triangleq E\left[\exp\left\{\boldsymbol{\tau}^T\mathbf{s}\right\}\right]$. Its first derivative (gradient) and second derivative (Hessian) with respect to $\boldsymbol{\tau}$ are denoted $\boldsymbol{\varphi}_{\mathbf{s}}(\boldsymbol{\tau})\triangleq\frac{\partial\varphi_{\mathbf{s}}(\boldsymbol{\tau})}{\partial\boldsymbol{\tau}}=E\left[\exp\left\{\boldsymbol{\tau}^T\mathbf{s}\right\}\cdot\mathbf{s}\right]\in\mathrm{R}^{\mathrm{K}\times 1}$, and $\boldsymbol{\Phi}_{\mathbf{s}}(\boldsymbol{\tau})\triangleq\frac{\partial^2\varphi_{\mathbf{s}}(\boldsymbol{\tau})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^T}=E\left[\mathbf{s}\mathbf{s}^T\cdot\exp\left\{\boldsymbol{\tau}^T\mathbf{s}\right\}\right]\in\mathrm{R}^{\mathrm{K}\times\mathrm{K}}$, respectively.

The log-GCF is defined as $\psi_{\mathbf{s}}(\boldsymbol{\tau})\triangleq\log\varphi_{\mathbf{s}}(\boldsymbol{\tau})$, and its gradient and Hessian are similarly defined and denoted $\boldsymbol{\psi}_{\mathbf{s}}(\boldsymbol{\tau})$ and $\boldsymbol{\Psi}_{\mathbf{s}}(\boldsymbol{\tau})$, respectively.

Now, assuming the model $\mathbf{x}=\mathbf{A}_0\mathbf{s}$, we have

$$\varphi_{\mathbf{x}}(\boldsymbol{\tau})=E\left[\exp\left(\boldsymbol{\tau}^T\mathbf{x}\right)\right]=E\left[\exp\left(\boldsymbol{\tau}^T\mathbf{A}_0\mathbf{s}\right)\right]=\varphi_{\mathbf{s}}\left(\mathbf{A}_0^T\boldsymbol{\tau}\right),$$

which implies that $\psi_{\mathbf{x}}(\boldsymbol{\tau})=\psi_{\mathbf{s}}\left(\mathbf{A}_0^T\boldsymbol{\tau}\right)$, and therefore that the Hessians $\boldsymbol{\psi}_{\mathbf{x}}(\boldsymbol{\tau})$ and $\boldsymbol{\psi}_{\mathbf{s}}(\boldsymbol{\tau})$ are related by

$$\boldsymbol{\Psi}_{\mathbf{x}}(\boldsymbol{\tau})=\mathbf{A}_0\boldsymbol{\Psi}_{\mathbf{s}}\left(\mathbf{A}_0^T\boldsymbol{\tau}\right)\mathbf{A}_0^T$$

[2] Generally, the processing-points can be complex-valued; However, to simplify the exposition we only consider real-valued processing points in here.

It is easy to observe (see [3] for details), that if the elements of $\mathbf{s}$ are statistically independent, then $\boldsymbol{\Psi}_{\mathbf{s}}(\boldsymbol{\tau})$ is a diagonal matrix for all $\boldsymbol{\tau}$. Therefore, any set of Hessian matrices of the log-GCF of $\mathbf{x}$ taken at several (M) processing-points, say $\boldsymbol{\Psi}_{\mathbf{x}}(\boldsymbol{\tau}_1),...,\boldsymbol{\Psi}_{\mathbf{x}}(\boldsymbol{\tau}_M)$, is jointly diagonalized by $\mathbf{A}_0^{-1}$. Therefore, these Hessian matrices can be used (under mild conditions, see [3]) for estimating the mixing matrix using (exact) joint diagonalization. In practice, however, these matrices are obviously unknown, and have to be estimated from the data.

It is shown in [3], that convenient and consistent estimation of the Hessian of the log-GCF can be formulated as "specially-weighted empirical covariance matrices",

$$\hat{\boldsymbol{\Psi}}_{\mathbf{x}}(\mathbf{x},\boldsymbol{\tau})=\left(\sum_{t=1}^{T}w_t\right)^{-1}\sum_{t=1}^{T}w_t\left(\mathbf{x}[t]-\bar{\mathbf{x}}\right)\left(\mathbf{x}[t]-\bar{\mathbf{x}}\right)^T, \quad (1)$$

where $w_t\triangleq\exp\left\{\boldsymbol{\tau}^T\mathbf{x}[t]\right\}\in\mathbb{R}$ can be regarded as weights, and $\bar{\mathbf{x}}\triangleq\left(\sum_{t=1}^{T}w_t\right)^{-1}\sum_{t=1}^{T}w_t\mathbf{x}[t]\in\mathbb{R}^{d\times 1}$ is a similarly-averaged empirical mean.

Note that for real-valued processing points $\boldsymbol{\tau}\in\mathbb{R}^{d\times 1}$, all the weights $w_t\in\mathbb{R}$ are positive, leading to positive-definite weighting, and to guaranteed positive semi-definite matrices $\hat{\boldsymbol{\Psi}}_{\mathbf{x}}(\mathbf{x},\boldsymbol{\tau})$. Also, for the particular case $\boldsymbol{\tau}=\mathbf{0}$, all weights equal one, and $\hat{\boldsymbol{\Psi}}_{\mathbf{x}}(\mathbf{x},\mathbf{0})$ is the sample covariance. Therefore, the sample covariance-matrix is often naturally included as one of the empirical Hessian matrices in the set of target-matrices for AJD, corresponding to one particular processing-point ($\boldsymbol{\tau}=\mathbf{0}$). The other empirical Hessians in the set correspond to other (off-origin) processing-points.

Unsurprisingly, the performance of CHESS depends on the selection of processing-points, and without prior knowledge regarding the observations' distributions, it is difficult to predict which processing points would yield better separation performance. Since the number of processing-points to be used (namely, the number of matrices in the AJD) is theoretically unrestricted, a natural tendency is to arbitrarily take as many processing-points as computationally affordable, hoping that at least some of these points would be "good". But then, if no weighting is used in the AJD process, there is a risk that the presence of (unknown) "bad" points in the group would result in degraded performance, relative to the performance that could have been attained if the only the (unknown) "good" points were taken. By estimating the covariance of the elements of these matrices (from the available data) and properly incorporating its inverse in the AJD process, the "good" points are automatically given "the upper-hand" in the combination (loosely speaking, of course).

2.2 WEDGE – WEighted Diagonalization using Gauss itErations

The WEDGE algorithm is developed in [5], where its use is described for the weighted AJD of estimated lagged correlation matrices – giving rise to an efficient implementation of WASOBI. Following is a brief outline of the algorithm.

WEDGE

1. Inputs:
   a. Set of target-matrices: $\hat{\mathbf{\Psi}}_{\mathbf{x}}[m] \in \mathbb{R}^{d\times d}, 1 \le m \le M$
   b. Set of weight matrices: $\mathbf{W}_{kl} \in \mathbb{R}^{M\times M}, 1 \le l < k \le d$
   c. An initial guess $\hat{\mathbf{V}}^{[0]} \in \mathbb{R}^{d\times d}$, e.g., $\hat{\mathbf{V}}^{[0]} = \left(\hat{\mathbf{\Psi}}_{\mathbf{x}}[1]\right)^{-\frac{1}{2}}$
2. $j \leftarrow 1$
3. Repeat 4-9 until convergence:
4. $\hat{\mathbf{\Psi}}_{\mathbf{s}}[m] = \hat{\mathbf{V}}^{[j-1]}\hat{\mathbf{\Psi}}_{\mathbf{x}}[m]\left(\hat{\mathbf{V}}^{[j-1]}\right)^{T}$ for $1 \le m \le M$.
5. $\hat{\mathbf{r}}_{k\ell} = \left[\left[\hat{\mathbf{\Psi}}_{\mathbf{s}}[1]\right]_{k,\ell} \quad \cdots \quad \left[\hat{\mathbf{\Psi}}_{\mathbf{s}}[M]\right]_{k,\ell}\right]^{T}$ for each $1 \le \ell \le k \le d$.
6. Set $\hat{\mathbf{A}}_{k,k} = 1$ for $1 \le k \le d$, whereas for $1 \le \ell \ne k \le d$ substitute the solutions of:

$$\begin{bmatrix}\hat{\mathbf{A}}_{k,\ell}\\ \hat{\mathbf{A}}_{\ell,k}\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{r}}_{\ell\ell}^{T}\mathbf{W}_{k\ell}\hat{\mathbf{r}}_{\ell\ell} & \hat{\mathbf{r}}_{kk}^{T}\mathbf{W}_{k\ell}\hat{\mathbf{r}}_{\ell\ell}\\ \hat{\mathbf{r}}_{kk}^{T}\mathbf{W}_{k\ell}\hat{\mathbf{r}}_{ll} & \hat{\mathbf{r}}_{kk}^{T}\mathbf{W}_{k\ell}\hat{\mathbf{r}}_{kk}\end{bmatrix}^{-1}\begin{bmatrix}\hat{\mathbf{r}}_{\ell\ell}^{T}\mathbf{W}_{k\ell}\hat{\mathbf{r}}_{k\ell}\\ \hat{\mathbf{r}}_{kk}^{T}\mathbf{W}_{k\ell}\hat{\mathbf{r}}_{k\ell}\end{bmatrix} \qquad (2)$$

7. Set $\hat{\mathbf{V}}^{[j]} = \left(\hat{\mathbf{A}}\right)^{-1}\hat{\mathbf{V}}^{[j-1]}$
8. Normalize the rows of $\hat{\mathbf{V}}^{[j]}$ such that $\left[\hat{\mathbf{V}}^{[j]}\hat{\mathbf{\Psi}}_{\mathbf{x}}[1]\left(\hat{\mathbf{V}}^{[j]}\right)^{T}\right]_{k,k} = 1 \qquad 1 \le k \le d$
9. $j \leftarrow j+1$
10. Outputs: The estimated demixing matrix $\hat{\mathbf{V}}$

(Note that $[\bullet]_{k,\ell}$ denotes the (k,ℓ)-th element of the enclosed matrix, whereas $\hat{\mathbf{r}}_{k\ell}$ and $\mathbf{W}_{k\ell}$ are the (k,ℓ)-th vector and weight-matrix, as defined above).

It was shown in [5] that equation (2) is the solution of a single Gauss iteration in the nonlinear WLS problem of weighted AJD (with the specified block-diagonal structure of the weight matrix), starting from an initial guess $\mathbf{A} = \mathbf{I}$. The algorithm consists of re-transforming the target-matrices (in step 7) after each iteration, such that the initial guess for the next iteration can again be taken as the identity matrix, leading to the conveniently decoupled solution of the Gauss iteration in (2).

The solution of the $d(d-1)/2$ sets of 2×2 equations in (2) has complexity $O\left(M^2 d^2\right)$, and this is also the complexity of each iteration.

3. WEIGHT MATRICES FOR CHESS

As already mentioned above, the optimal weight matrix (in the sense of minimum mean square error in the estimation of elements of the demixing matrix $\hat{\mathbf{V}}$) for the weighted AJD process is (assuming small errors in the estimated target-matrices) the inverse of the covariance matrix of all off-diagonal elements in the target-matrices (properly stacked in a single vector). When the signals are nearly separated, this covariance matrix possesses the appealing property of being block diagonal, with $d(d-1)/2$ blocks across the diagonal, each of size $M \times M$, defined as:

$$\mathbf{W}_{k\ell} = Cov^{-1}\left(\hat{\mathbf{r}}_{k\ell}\right) \in \mathbb{R}^{M\times M},\ 1 \le \ell < k \le d, \qquad (3)$$

where $\hat{\mathbf{r}}_{k\ell}$ has been defined in the description of the WEDGE algorithm above, on step 5. This was rigorously shown to be the case for the SOBI target-matrices (in [5]) and for the JADE target-matrices (in [6]).

An explicit, exact analytical expression for the empirical Hessians' covariance matrices is difficult to obtain, but an empirical estimate of this covariance from the data can be obtained by dividing the observation interval into blocks and estimating the covariance (over blocks) of the empirical Hessians obtained in each block. The weight matrices $\mathbf{W}_{kl}$ are then set to the inverses of these matrices.

3.1 Error-covariance Estimation:

1. Inputs: the observed data $\hat{\mathbf{x}}[t], 1 \le t \le T$ and M arbitrarily-selected "processing-points" $\boldsymbol{\tau}_1, ..., \boldsymbol{\tau}_M$.
2. Partition the observation interval $1 \le t \le T$ into P blocks of equal lengths.
3. For each block $1 \le p \le P$, repeat steps 4, 5:
4. Estimate $\hat{\mathbf{\Psi}}_{\mathbf{x}}^{(p)}[m], 1 \le m \le M$ using equation (1).
5. Vectorize:
   $\hat{\mathbf{r}}_{k\ell}^{(p)} \triangleq \left[\left[\hat{\mathbf{\Psi}}_{\mathbf{x}}^{(p)}[1]\right]_{k,\ell} \quad \cdots \quad \left[\hat{\mathbf{\Psi}}_{\mathbf{x}}^{(p)}[M]\right]_{k,\ell}\right]^{T}$
   $1 \le l < k \le d$
7. For each $1 \le l < k \le d$, obtain the covariance estimate:
   $\hat{\mathbf{C}}_{k\ell} = \frac{1}{P-1}\sum_{p=1}^{P}\left(\hat{\mathbf{r}}_{k\ell}^{(p)} - \bar{\hat{\mathbf{r}}}_{k\ell}\right)\left(\hat{\mathbf{r}}_{k\ell}^{(p)} - \bar{\hat{\mathbf{r}}}_{k\ell}\right)^{T}$,
   where $\bar{\hat{\mathbf{r}}}_{k\ell} \triangleq \frac{1}{P}\sum_{p=1}^{P}\hat{\mathbf{r}}_{k\ell}^{(p)}$
8. Outputs: The weight matrices
   $\mathbf{W}_{k\ell} = \mathbf{C}_{k\ell}^{-1}$, $1 \le l < k \le d$.

3.2 The iterative reweighting algorithm

The efficiency benefit of WEDGE is achieved when the signals are nearly separated. However, usually the given mixture is far from this state, and a preliminary separation step should be applied first. Therefore the following iterative

algorithm is proposed. In our simulation examples (shown in Section 4), convergence was attained within 2-3 iterations.

Weighted CHESS:

1. Input: the observed data $\mathbf{x}[t], 1 \le t \le T$ and M arbitrarily-selected "processing-points" $\boldsymbol{\tau}_1, ..., \boldsymbol{\tau}_M$.
2. $\hat{\mathbf{V}}_0 \leftarrow$ an initial guess for the separation matrix, provided by some consistent BSS algorithm – e.g., unweighted CHESS, set $\hat{\mathbf{x}}_0[t] = \mathbf{x}[t], 1 \le t \le T$.
3. $j \leftarrow 1$
4. Repeat steps 5-9 until convergence:
5. $\hat{\mathbf{x}}_j[t] \leftarrow \hat{\mathbf{V}}_{j-1}\hat{\mathbf{x}}_{j-1}[t]$, $1 \le t \le T$
6. (Re-)generate the target-matrices: $\hat{\boldsymbol{\Psi}}_{\mathbf{x}}[m] \leftarrow \hat{\boldsymbol{\Psi}}_{\mathbf{x}}\left(\hat{\mathbf{x}}_j, \boldsymbol{\tau}_m\right)$ using (1) for $1 \le m \le M$;
7. (Re-)estimate the weight matrices: $\{\mathbf{W}_{k\ell}\}_{k>\ell=1}^{d} \leftarrow$ Weight-Estimation with $\hat{\mathbf{x}}_j[t]$;
8. $\hat{\mathbf{V}}_j \leftarrow$ WEDGE with $\left\{\hat{\boldsymbol{\Psi}}_{\mathbf{x}}[m]\right\}_{m=1}^{M}, \{\mathbf{W}_{k\ell}\}_{k>\ell=1}^{d}$;
9. $j \leftarrow j+1$
10. Upon convergence ($j = J$):

Output: $\hat{\mathbf{V}} = \hat{\mathbf{V}}_J \hat{\mathbf{V}}_{J-1} \cdots \hat{\mathbf{V}}_0$, the estimated separated signals are $\hat{\mathbf{s}}[t] = \hat{\mathbf{x}}_J[t]$, $1 \le t \le T$

4. SIMULATION

As mentioned earlier, the relative computational efficiency of our proposed reweighted CHESS algorithm is based on the assumed block-diagonality of the covariance matrix of the concatenated vector of the sample Hessians of the log-GCF of the observation in a "near-separation" condition. Evidently, such (near) block-diagonality implies that $\hat{\mathbf{r}}_{k\ell}, \hat{\mathbf{r}}_{k'\ell'}$ are (nearly) uncorrelated for all $(k,\ell) \ne (k',\ell')$. A rigorous proof of that property is too long to be included in this contribution, but we illustrate this block-diagonality in Figure 1 for a particular case with $d = 3, M = 5$. The figure shows the intensity of absolute values of the empirical covariance matrix of the concatenated vector $\mathbf{r} = \left[\hat{\mathbf{r}}_{12}^T \quad \hat{\mathbf{r}}_{13}^T \quad \hat{\mathbf{r}}_{23}^T\right]^T$. The expected block-diagonal structure of $d(d-1)/2 = 3$ block of dimension $M \times M = 5 \times 5$ is clearly observed.

To capture the essence and effectiveness of the proposed weighted approach, we consider a simple scenario of $d = 2$ sources, each with a zero-mean unit-variance distribution. We present the results of five experiments:

1. Unweighted CHESS with $M = 8$ processing-points (the selected processing-points are shown and marked with 'o' in Figure 2);
2. Unweighted CHESS with $M = 8 + 7$ processing-points, which are the same processing-points used in the first experiment, with additional 7 processing points used in the second experiment (shown and marked by '*' in Figure 2);
3. The proposed weighted version of CHESS with the same processing-points as in Experiment 1;
4. The proposed weighted version of CHESS with the same processing-points as in Experiment 2;
5. For reference we also provide the performance of JADE.

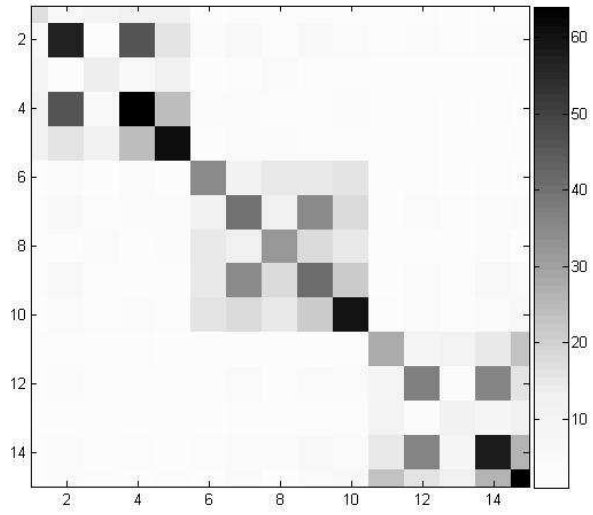

Figure 1: Visualization of the covariance matrix of the off-diagonal elements of the sample-Hessian of the observations' log-GCF

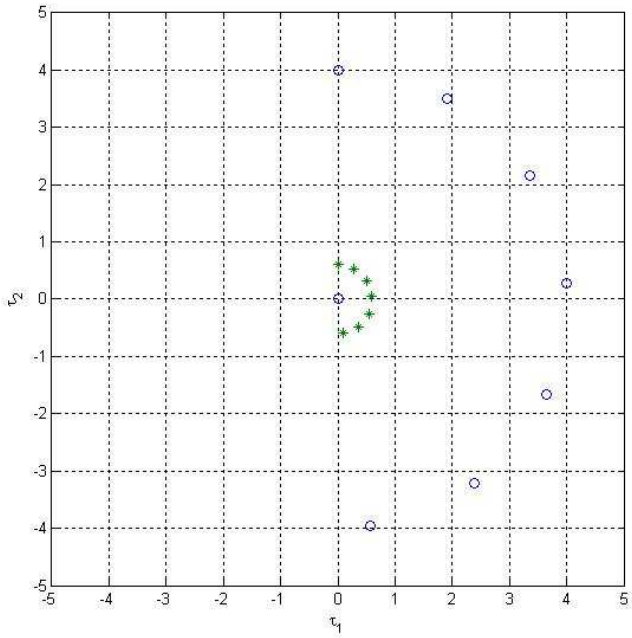

Figure 2 – The processing-points constellations used in Experiments 1-4

The results are shown in Figure 3 vs. the observation length T, in terms of the attained mean Interference to Source Ratio (ISR) of the two channels, averaged along 400 independent trials. In all the experiments, the various algorithms were applied to the same data in each trial.

As evident in Figure 3, while the initial choice of the processing-points for ordinary (unweighted) CHESS in Experiment 1 slightly outperforms JADE, the performance is degraded when the additional processing-points are added in Experiment 2. Initially, this may possibly appear as somewhat counter-intuitive: although more target-matrices are added to the joint diagonalization process, the performance degrades. This is exactly where the optimal weighting comes into play: obviously, the addition of data (target-matrices) is guaranteed not to degrade (or even to improve) the performance, only if such additional data is properly weighted into the estimation process. If the additional data is "worse" than the original data, then attributing uniform weight to all the data would evidently result in loss of performance.

Indeed, from the results of Experiment 4 we observe that the proper weighting of Hessians from all of the processing-points attains a significant performance improvement, not only with respect to the performance attained by the unweighted use of Hessians from the original processing-points in Experiment 1, but also with respect to the optimally-weighted use of these Hessians alone in Experiment 3. In other words, while in an unweighted framework the additional processing-points degrade the performance, in a properly-weighted framework the same additional data can improve the performance.

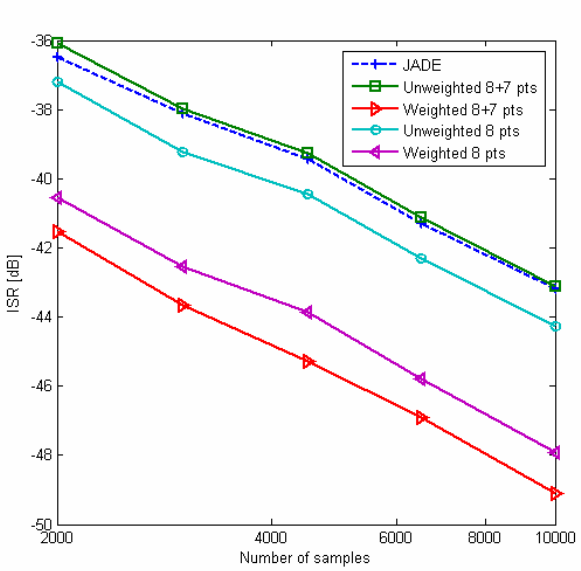

Figure 3: The Result of the five experiments: ISR [dB] vs. the observation length T .

5. CONCLUSIONS

We considered optimal weighting of the sample Hessians of the observations' log-GCF in the AJD process in CHESS. Generally, such optimal weighting involves the estimation of the covariance matrix of all the elements of the sample Hessians, followed by inversion of this covariance matrix in order to obtain the estimated optimal weight matrix. However, when the sources are nearly-separated, this covariance matrix becomes block-diagonal, such that its estimation is conveniently reduced to estimation of the particular blocks; Likewise, the optimal weight matrix is reduced to a block-diagonal matrix, whose blocks are obtained as the inverses of the respective blocks in the covariance matrix. Moreover, such block-diagonal weighting is conveniently applied to the AJD process using the recently proposed WEDGE algorithm.

Thus, our proposed approach first applies an initial separation stage, and then further refines the separation by iteratively applying the aforementioned weighting scheme to the nearly-separated observations.

Using simulation results we demonstrated the ability of the proper weighting to better exploit any given choice of processing-points: When optimal weighting is used, the addition of processing-points cannot degrade the mean performance, since, in the "worst case", a "bad" processing-point would be naturally out-weighted by the algorithm, and will not cause any damage. Of course, with any finite data-length, the estimated weights might not be the true optimal weights, and therefore in practice such monotonic improvement (with an increase in the number of processing points) is generally not guaranteed. With a restricted number of processing-points, the CHESS results are still quite sensitive to the selection of the processing points. A future direction of this research is to find optimal choice and number of processing points, depending solely on the observed data (and with a reasonable computational complexity).

6. BIBLIOGRAPHY

[1] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals", *IEE Proc.-F*, vol. 140(6), pp. 362-370, 1993.

[2] 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.

[3] A. Yeredor, "Blind source separation via the second characteristic function", *Signal Processing*, Vol. 80(5), pp. 897–902, 2000.

[4] 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.

[5] P. Tichavský and A. Yeredor, "Fast Approximate Joint Diagonalization Incorporating Weight Matrices", *IEEE Trans. on Signal processing*, 57(3), pp. 878-891, 2009.

[6] A. Smekhov and A. Yeredor, "Optimization of JADE Using a Novel Optimally Weighted Joint Diagonalization Approach", in *Proc. EUSIPCO 2004*, Vienna, Austria, Sep. 2004.

[7] P. Tichavsky, E. Doron, A. Yeredor and J. Nielsen, "A Computationally Affordable Implementation of an Asymptotically Optimal BSS Algorithm for AR Sources", in *Proc. EUSIPCO 2006*, Florence, Italy, Sep. 2006.

[8] E. Eidinger and A. Yeredor, "Blind Source Separation via the Second Characteristic Function with Asymptotically Optimal Weighting", in *Proc., 23rd IEEE Israel Convention (IEEEI2004)*, Herzliya, Israel, Sep., 2004.