CRITERIA FOR THE SIMULTANEOUS BLIND EXTRACTION OF ARBITRARY GROUPS OF SOURCES

Sergio Cruces

Signal Processing Group
Camino Descubrimientos,
41092-Seville, Spain.
sergio@cica.es

Andrzej Cichocki, Shun-ichi Amari

Brain Science Institute
RIKEN, 2-1 Hirosawa, Wako-shi,
Saitama, 351-0198 Japan.
{cia,amari}@brain.riken.go.jp

ABSTRACT

This paper reviews several existing contrast functions for blind source extraction proposed in the areas of Projection Pursuit and Independent Component Analysis, in order to extend them to allow the simultaneous blind extraction of an arbitrary number of sources which is specified by user. Using these criteria a novel form of Amari's extraction algorithm has been derived. The necessary and sufficient asymptotical stability conditions that we obtain for this algorithm help us to develop step-sizes that result in a fast convergence. Finally, we exhibit some exemplary simulations that validate our theoretical results and illustrate the excellent performance of the presented algorithms.

1. INTRODUCTION

Blind Source Separation (BSS) is the problem of recovering mutually independent unobserved signals (sources) from their linear mixture. Although this problem has recently attracted a lot of interest because of its wide number of applications in diverse fields, BSS can be very computationally demanding if the number of source signals is large (say, of order of 100 or more). In particular, this is the case in biomedical signal processing applications such as EEG/MEG data processing where the number of sensors can be larger than 120 and it is desired to extract only some 'interesting' sparse sources. Fortunately, sequential Blind Source Extraction (BSE) overcomes somewhat this difficulty. The BSE problem considers the case where only a small subset of sources have to be recovered from a large number of sensor signals.

The combined use of BSE and deflation to solve the BSS problem was originally proposed in [1] and further explored in [2, 3, 4]. Related research has been done in the area of Exploratory Projection Pursuit, with the aim of obtaining low-dimensional informative views of high-dimensional data, and several criteria and algorithms for BSE has been much earlier proposed (see [5] for a good survey of these techniques).

The main limitation of existing BSE algorithms is that most of them can only recover the sources sequentially one by one. The principal reason for this behavior is the necessity to avoiding the possibility of obtaining, at the outputs, the sources replicated. Only few algorithms have been proposed till now which enable to extract simultaneously an arbitrary group of 'interesting' sources (from 1 till $E \leq N$ where N is the total number of source signals and E is the number of sources extracted simultaneously, specified by user) being, up to our knowledge, the only exceptions [6, 7].

In this paper we present a straightforward technique that allow us to extend some of the classical criteria for blind source separation and extraction to the case of the simultaneous blind source extraction of an arbitrary subgroup of E ($1 \leq E \leq N$) sources.

The structure of the paper is as follows. Section 2 will specify the considered signal model and notation. Section 3 motivates the difficulty of blind source extraction and presents the extension of several existing criteria to allow the simultaneous extraction of a subset of the sources. In section 4 we propose a special form of Amari's algorithm which uses the Stiefel manifolds to satisfy the discussed constrained optimization criteria. In section 5, we present practical bounds for the algorithm step-size derived from the asymptotical stability analysis. Section 6 presents exemplary simulation results, and finally, section 7 presents the conclusions.

2. SIGNAL MODEL AND NOTATION

Let us consider the standard linear mixing model of N unknown statistically independent source signals $\mathbf{s} = [s_1, \cdots, s_N]^T$ that are linearly combined by the memoryless system described by a mixing matrix $\mathbf{A}$ to give the vector of observations

$$\mathbf{x} = \mathbf{A}\mathbf{s} \,. \tag{1}$$

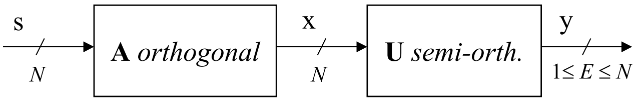


Fig. 1. Considered signal model for simultaneous blind source extraction.

Without loss of generality we assume that the sources are zero mean with unit variance random variables and the unknown mixing matrix $\mathbf{A}$ is orthogonal. Note, that the orthogonality of the mixing matrix ($\mathbf{A}\mathbf{A}^T = \mathbf{I}_N$) can be always enforced by simply performing pre-whitening on the original observations. For noisy data the robust prewhitening or orthogonalization can be employed.

In order to extract $E < N$ sources, the observations will be further processed by an $E \times N$ semi-orthogonal separating matrix $\mathbf{U}$, satisfying $\mathbf{U}\mathbf{U}^T = \mathbf{I}_E$, which yields to the outputs vector (or estimated sources)

$$\mathbf{y} = \mathbf{U}\mathbf{x} = \mathbf{G}\mathbf{s} \; , \tag{2}$$

where $\mathbf{G} = \mathbf{U}\mathbf{A}$ will denote the semi-orthogonal $E \times N$ global transfer matrix from the sources to the outputs. The semi-orthogonality of the global transfer system will be important for preserving the spatial decorrelation of the outputs vector since $Cov(\mathbf{Y}) = E[(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{Y} - E[\mathbf{Y}])^T] = \mathbf{G}\mathbf{G}^T = \mathbf{I}_E$.

Along the paper we will follow this notation. Usually we will work with normalized random variables, i.e., those with zero mean and unit variance. According to the usual criteria random variables will be denoted in capital letters, whereas the samples of these variables will appear in lowercase letters. We will reserve the symbol $\mathcal{G}$ to specify when a random variable is Gaussian distributed. C_Y^r will denote the r-th order auto-cumulant of the random variable, i.e., $C_Y^r = Cum(Y \times r)$. The differential entropy of the random variable Y will be denoted by

$$h(Y) = -\int p_Y(y) log p_Y(y) dy \tag{3}$$

and the Kullback-Leibler divergence between two densities $p_F(y)$ and $p_G(y)$ will be referred by

$$KL(p_F || p_G) = \int p_F(y) log \frac{p_F(y)}{p_G(y)} dy \tag{4}$$

3. CRITERIA FOR THE SIMULTANEOUS BLIND SOURCE EXTRACTION

According to the Darmois-Skitovich theorem [8], the most natural criterium for the BSE of E sources is the minimization of the mutual information of the outputs [9, 10]

$$I(Y_1, \ldots, Y_E) \;\; = \;\; -h(Y_1, \ldots, Y_E) + \sum_{i=1}^{E} h(Y_i) \tag{5}$$

since the independence and the non-Gaussianity of the desired sources are the key assumptions in this problem.

In Blind Source Separation ($E = N$) this was the starting approach for several interesting algorithms. The main difficulty in the application of this criterium consist in the necessity of the estimation of the joint entropy of the outputs since this will also involve the estimation of their joint p.d.f., a non-trivial task that would require an extensive amount of data and computational resources. However, under the spatial decorrelation constraint of the outputs, the joint entropy $h(Y_1, \ldots, Y_N)$ is kept constant and the criterium turns implementable. What it remains to do is to perform the minimization of the sum of the marginal entropies and this has been done in two different ways since each marginal entropy can be approximated using the truncated Edgeword [9] or the Gram-Charlier [11] expansions of the marginal p.d.f.s of the outputs.

Unfortunately, the same trick can not be used for the blind extraction with $E < N$. On the contrary to the Blind Source Separation case, the joint entropy $h(Y_1, \ldots, Y_E)$ now will depend on the subspace spanned by rows of the extracting system $\mathbf{U}$ being, therefore, nonconstant. The reason for this behavior is that the extracting system allows us to extract different sets of E sources which, at the same time, have different joint entropies.

Once again, in order to overcome this difficulty, we should avoid the explicit use of joint densities. With this aim, Huber [5] suggested to find a functional $\psi(\cdot)$ that maps each p.d.f. of a normalized random variable Y_i to a real index $\psi(Y_i)$ that satisfies the following *properties*:

1. $\psi(\cdot)$ is affine invariant.

2. $\psi(\cdot) \geq 0$ and the minimum value of the index ($\psi(Y_i) = 0$) is obtained when $p_{Y_i} = p_{\mathcal{G}}$, i.e., when the r.v. follows a Gaussian distribution.

3. $\psi(\cdot)$ is convex (strictly convex) with respect to the linear combinations of the independent sources (of the independent sources for which $\psi(S_j) \neq 0$), in such a way that, if $Y_i = \sum_{j=1}^{N} G_{ij} S_j$, then

$$\psi\left(Y_i\right) \leq \sum_{j=1}^{N} |G_{ij}|^2 \psi(S_j) \tag{6}$$

where G_{ij} are the elements of the semi-orthogonal matrix $\mathbf{G}$ and $S_j, j = 1, \ldots, N$, are independent and normalized random variables.

These three properties are close to those given by Donoho [12] and Comon [9] when they defined the idea of contrast functions. Many functionals that satisfy the previous properties has been proposed in the literature [5, 13, 4]. Next, will briefly review and summarize some of these criteria:

Minimum Entropy: The minimum Entropy criteria first introduced by Wiggins [14, 12] in single channel Blind Deconvolution and later proposed by Huber [5] in Projection Pursuit is given by the maximization of

$$\psi_{ME}(Y_1) = KL(p_{Y_1}(y_1)||p_{\mathcal{G}_1}(y_1)) \quad (7)$$

$$= \frac{1}{2}\log(2\pi e) - h(Y_1) \quad (8)$$

where Y_1 and $\mathcal{G}_1$ are, respectively, non-Gaussian and Gaussian normalized random variables. Another form of this criterium has been proposed by Friedman in [13]. If we perform a transformation of the random variable Y_i to another r.v. with the following bounded support $Z_1 = 2\Phi(Y_1) - 1 \in [-1, 1]$ (were $\Phi(Y_1)$ denotes the standard Gaussian c.d.f.), then an alternative form of the Minimum Entropy index is given by

$$\psi_{ME}(Y_1) = KL(p_{Z_1}(z_1)||p_{\mathcal{U}_1}(z_1)) \quad (9)$$

where $\mathcal{U}_1$ is an uniform random variable defined in the interval [-1,1].

Maximum Likelihood: Recently, Amari *et al.* proposed in [15, 6] another index for BSE closely related with the Maximum Likelihood approach. Let S_1 be the normalized source which, being part of the mixture, has the smaller differential entropy. Then,

$$\psi_{ML}(Y_1) = \frac{1}{2}\log(2\pi e) + E[\log p_{S_1}(y_1)] \quad (10)$$

and it is ready seen that $\psi_{ML}(Y_1) \leq \psi_{ME}(Y_1)$.

Cumulants based index: This index has a long history, different authors has proposed it in many different ways and forms [14, 9, 1, 4, 2, 16]. One general form of this index, for normalized r.v., is given by

$$\psi_{Cum}(Y_1) = \sum_{r>2} \omega'_r \cdot |C^r_{Y_1}|^{\alpha_r} \quad (11)$$

where $\alpha_r \geq 1$, $|C^r_{Y_1}|$ denotes the modulo of the r-th order auto-cumulant and $\omega'_r = \frac{\omega_r}{r\alpha_r}$ are scaled or normalized non-negative weighting factors. Note that the low-order cumulants (for r=1 and r=2) are rather excluded from the index since these are kept fixed by the normalization constraint of the random variable.

It is well known that the previous indices measure the degree of non-Gaussianity or the amount of structure that is present in the outputs and that, from property 3, they have their maxima at the extraction of one of the independent sources. Thus, the blind extraction of one of the sources is obtained solving the following constrained maximization problem

$$\max_{\mathbf{U}} \psi(Y_1) \text{ subject to } Cov(Y_1) = 1 \quad (12)$$

whereas, the blind source separation of the whole set of sources is obtained maximizing

$$\max_{\mathbf{U}} \sum_{i=1}^{N} \psi(Y_i) \text{ subject to } Cov(\mathbf{Y}) = \mathbf{I}_N \quad (13)$$

However, one can observe that there is a theoretical gap between the extraction of one source and the extraction of the whole set of sources. The next theorem will establish some links between both approaches.

Theorem 1 *Given a functional $\psi(\cdot)$ that satisfies satisfies properties 1-3, if the sources can be ordered by decreasing value of this functional as*

$$\psi(S_1) \geq \ldots \geq \psi(S_E) > \psi(S_{E+1}) \geq \ldots \geq \psi(S_N) \quad (14)$$

and if $\psi(S_E) > 0$, then the following objective function

$$\Psi(\mathbf{Y}) = \sum_{i=1}^{E} \psi(Y_i) \textit{ subject to } Cov(\mathbf{Y}) = \mathbf{I}_E \quad (15)$$

will be a contrast function whose global maxima correspond to the extraction of the first E sources of the mixture, i.e., at this maxima $\mathbf{Y} = [S_1, \ldots, S_E]^T$ up to an arbitrary reordering or permutation between them.

Proof: From property 3 we have that

$$\sum_{i=1}^{E} \psi(Y_i) \leq \sum_{j=1}^{N} \psi(S_j) \sum_{i=1}^{E} |G_{ij}|^2 \quad (16)$$

$$= trace\{\mathbf{G}\mathbf{\Lambda}\mathbf{G}^T\} \quad (17)$$

where $\mathbf{\Lambda}$ is a diagonal matrix which elements are $\Lambda_{jj} = \psi(S_j)$ for $j = 1, \ldots, N$.

But the decorrelation constraint for the outputs ($Cov(\mathbf{Y}) = \mathbf{I}_E$) is tantamount to the semi-orthogonality of the global transfer matrix $\mathbf{G}$. From the application of the Poincaré's separation theorem of matrix algebra, and according to the sources ordering (14), the eigenvalues $\sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_E$ of $\mathbf{G\Lambda G}^T$ satisfy

$$\psi(S_{N-E+j}) \leq \sigma_j \leq \psi(S_j) \quad (18)$$

Thus, the maximum of (17) subject to the semi-orthogonality of $\mathbf{G}$ is

$$\max_{\mathbf{G}\mathbf{G}^T=\mathbf{I}_E} \text{trace}\{\mathbf{G}\Lambda\mathbf{G}^T\} = \sum_{j=1}^{E} \psi(S_j) \quad (19)$$

and the bound is attained only for that matrices $\mathbf{G}$ of which E rows consist in orthogonal vectors that span the same subspace of the eigenvectors associated with the E largest eigenvalues of Λ, which enforces that $G_{ij} = 0 \; \forall j > E$.

From the strict convexity of $\psi(\cdot)$, a necessary condition for the equality between (16) and (17) is that $G_{ij} = \{\pm 1 \text{ or } 0\}$ whenever $\psi(S_j) \neq 0$. Since, by assumption, $\psi(S_j) \neq 0 \ \forall j \leq E$, from the semi-orthogonality of $\mathbf{G}$ we can see that the necessary and sufficient condition for the equality between (16) and (19) is that the matrix $\mathbf{G}$ can be reduced by row permutations to the form $[\mathbf{I_E}, \mathbf{0}]$, i.e., $\mathbf{G}$ is the extraction matrix of the first E sources. □

4. EXTENDED AMARI'S ALGORITHM

A particularly simple and useful method to maximize the generic contrast function $\Psi(\mathbf{Y})$ is to use the natural Riemannian gradient ascent in the Stiefel manifold of semi-orthogonal matrices, which is given by

$$\tilde{\nabla}_{\mathbf{U}}\Psi = \nabla_{\mathbf{U}}\Psi - \mathbf{U}(\nabla_{\mathbf{U}}\Psi)^T\mathbf{U} \tag{20}$$

This leads to the algorithm proposed by Amari in [15] for blind source extraction

$$\mathbf{U}^{(k+1)} = \mathbf{U}^{(k)} - \mu\left(\mathbf{R}_{\varphi,x}^{(k)} - \mathbf{R}_{y,\varphi}^{(k)}\mathbf{U}^{(k)}\right) \tag{21}$$

where $\mathbf{R}_{\varphi,x}^{(k)} = E[\varphi(\mathbf{y})\,\mathbf{x}^T]$ and $\varphi(\mathbf{y}) = [-\frac{d\Psi}{dy_1}, \ldots, -\frac{d\Psi}{dy_E}]^T$.

The exact expressions of $\varphi(\mathbf{y})$ depend on the selected criteria. When $\Psi(\cdot) = \Psi_{ME}(\cdot)$ approximations to these derivatives can be found in [9] and in [11]. When using the cumulants based contrast $\Psi(\cdot) = \Psi_{Cum}(\cdot)$ the algorithm takes the novel form

$$\begin{aligned}\mathbf{U}^{(k+1)} &= \mathbf{U}^{(k)} + \mu\cdot \\ &\sum_{r>2}\omega_r\left(\mathbf{D}_y^r\mathbf{C}_{y,x}^{r-1,1} - \mathbf{C}_{y,y}^{1,r-1}\mathbf{D}_y^r\mathbf{U}^{(k)}\right)\end{aligned} \tag{22}$$

where $\mathbf{D}_y^r$ is the diagonal matrix with entries

$$[\mathbf{D}_y^r]_{ii} = \text{sign}(C_{y_i}^r)\cdot|C_{y_i}^r|^{(\alpha_r-1)} \tag{23}$$

and $\mathbf{C}_{y,y}^{1,r-1}$ is the r^{th}-order cross-cumulant matrix with elements $[\mathbf{C}_{y,y}^{1,r-1}]_{ij} = Cum(y_i[n], y_j[n]\times(r-1))$.

The advantage of the above approach is that, for algorithm (22), it is possible to guarantee the identifiability of the source extraction solutions. Nevertheless, there is no guarantee that we achieve always the global maximum since the gradient algorithm can be trapped in the local maxima corresponding to other extracting solutions.

Extensive simulation experiments show that it is usually sufficient to repeat the extraction procedure 2 or 3 times with deflation procedure to obtain all desired signals, which are, in our case, those with the largest index $\psi(Y_i)$ among all possible estimated sources. Although these index measures, in general, the departure from the Gaussianity of the sources, we still have some control in order to favor extraction of source signals with specific stochastic properties ordering through the proper selection of the involved cumulants orders r and the factors ω_r and α_r. For instance, if the sources of our interest have asymmetric distributions we can favor their extraction in first place by weighting more in the index (11) the skewness and other cumulants of odd order.

5. ASYMPTOTIC CONVERGENCE

Although a gradient algorithm with a proper initialization and a sufficient small step-size should have no problems to converge to the maxima of a given contrast, it is important to establish possibly large step sizes (learning rates) in order to ensure a high convergence rate and simultaneously guarantee the stability of the algorithm. Some bounds for the learning rate can be obtained from the asymptotical stability analysis of the algorithms.

In this section, we will adopt the following notation: in order to eliminate the permutation ambiguity in the extracted solution we define a vector of extracted sources $\mathbf{s}_E = [s_1, \cdots, s_E]^T$ which shares the same ordering of the outputs, $\varphi_i = [\varphi(\mathbf{s}_E)]_i$ is a non-linear function that acts component-wise on the extracted sources, $k_i = E\left[\frac{\partial\varphi_i}{\partial s_i} - s_i\varphi_i\right]$ is a variable (originally defined by Cardoso and Laheld in [17]) that will have an important role in the control of the stability of the algorithm.

The next two theorems presents the obtained stability results.

Theorem 2 *Assuming that the mixing system is orthogonal, the necessary and sufficient asymptotic stability conditions of Amari's algorithm, given by equation (21), to converge to the extraction solution are*

$$0 < \mu < \frac{2}{k_i} \qquad \textit{if } E = 1 \tag{24}$$

$$0 < \mu < \min\left\{\frac{2}{k_i+k_j}, \frac{2}{k_i}\right\} \qquad \textit{if } 1 < E < N \tag{25}$$

$$0 < \mu < \frac{2}{k_i+k_j} \qquad \textit{if } E = N \tag{26}$$

for all $i, j|_{i\neq j} = 1, \ldots, E$.

Due to the lack of space we will skip here the proof. However, it is interesting to observe that since Amari's algorithm takes the special form of the EASI algorithm (proposed by Cardoso and Laheld in [17]) in the particular case of blind source separation ($E = N$), the final condition for blind separation (26) is a simple extension of the local stability condition of the EASI algorithm [17]

$$k_i + k_j > 0 \quad \text{for all } 1 \leq i < j \leq N. \tag{27}$$

Theorem 3 *Assuming that the mixing system is orthogonal, the necessary and sufficient asymptotic stability conditions for the extended Amari's algorithm, of equation (22), to* *converge to the extraction solution, are also given by conditions (24)-(26), having, for this special case,*

$$k_i = \rho(s_i) = \sum_{r>2} \omega_r \cdot |C^r_{s_i}|^{\alpha_r} \tag{28}$$

Note that, since $\rho(s_i)$ is non-negative, the stability conditions only depend on the magnitude of the step-size. For this extended algorithm it is possible to bound the slowest convergence mode and determine a robust, but also close to the locally optimal, learning step size which results in a fast convergence. This is given by

$$\mu^{(k)} = \frac{2}{3\max_i \rho(y_i)} \tag{29}$$

The more similar be the $\rho(\cdot)$ factors for the sources, the faster it will be the convergence of algorithm (22) for this step size.

6. SIMULATIONS

With the purpose of an easy graphical representation of the results we will consider nine independent source images that are linearly combined by a random mixing matrix. These images have different kurtosis given by

$$\mathbf{C}^4_s = [\,4.2,\ 3.5,\ -2,\ 1.9,\ -1.6,\ -1,\ -1,\ 0.05,\ 0.01]$$

As we can observe that two of them are very close to Gaussian noise. After obtaining the observations, shown in figure 2, we perform prewhitening in order to decorrelate them.

We applied a batch version of Amari's algorithm which uses the contrast based on fourth order cumulants, i.e., equation (22) with $r = 4$, $\alpha_r = 1$ and $\omega_4 = 1$. Then, we run the extraction algorithm with $E = 3$ and with the adaptive step-size

$$\mu^{(k)} = \frac{2}{3\max_i |C^4_{y_i}|} \tag{30}$$

After 16 iterations the algorithm converged to the first three extracted sources shown in figure 3-a). Then, if these are not the sources of interest we can remove the contribution of these sources from the observation and perform a new extraction. The second extraction was obtained after 19 iterations to the three sources of figure 3-b) and, finally, the third extraction converged after 22 iterations to the three sources of figure 3-c). We can observe how, in agreement with the presented results, the algorithm favors the extraction of the sources with greater structure in the first attempts whereas the sources closer to being Gaussian or those with greater uncertainty are usually relegated to the last extractions.

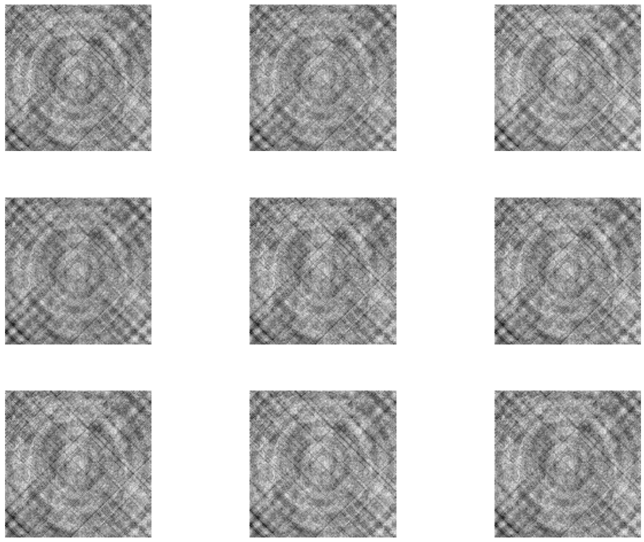


Fig. 2. Images of the observations before prewhitening.

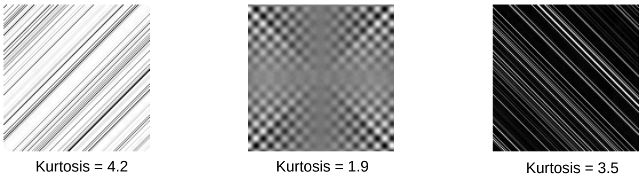
7. CONCLUSIONS

In this paper we have presented an unified extension of several existing optimization criteria for blind source separation/extraction to the simultaneous blind extraction of an arbitrary number of sources $E \leq N$. There are at least several reasons that justify the usefulness of the proposed approach including: simultaneous BSE has a lower computational burden than BSS, it is suitable for that applications where the deflation procedure should be avoided, and improves the robustness of the extraction in comparison to the sequential approach were the errors can be accumulated and propagated, moreover, applying global optimization or using only several trials we can extract sources with desired stochastic properties (i.e. those with the largest index $\psi(Y_i)$). An extended form of Amari's algorithm has been derived by applying natural gradient in the Stiefel manifold to a cumulants based contrast function. Furthermore, we have obtained the local stability conditions of this algorithm and established a learning rate that provides fast convergence. Finally, we have demonstrated by extensive computer simulations the validity of theoretical results and good performance of the proposed algorithm.

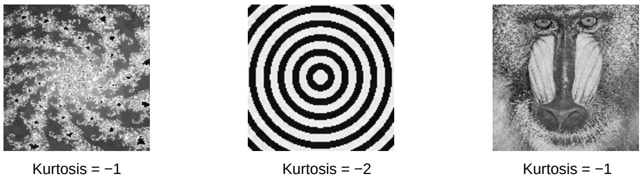
8. REFERENCES

[1] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, vol. 45, pp. 59–83, 1995.

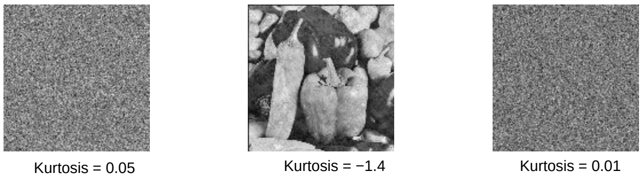
[2] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483–1492, 1997.



a) 1^{st} extraction: 16 iterations, $\Psi_{Cum}(\mathbf{y}) = 9.6$.



b) 2^{nd} extraction: 19 iterations, $\Psi_{Cum}(\mathbf{y}) = 4$.



c) 3^{rd} extraction: 22 iterations, $\Psi_{Cum}(\mathbf{y}) = 1.46$.

Fig. 3. Extraction of the nine sources in groups of three (E=3). After each extraction a deflation procedure has been applied.

[3] A. Cichocki, R. Thawonmas and S. Amari, "Sequential blind signal extraction in order specified by stochastic properties," *Electronics Letters*, vol. 33, no. 1, pp. 64–65, 1997.

[4] M. Girolami and C. Fyfe, *Negentropy and kurtosis as projection pursuit indices provide generalized ICA algorithms*, pp. 752–763, Boston, MA: MIT Press, 1996.

[5] Peter J. Huber, "Projection pursuit," *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, 1985.

[6] S. Amari, A. Cichocki, and H.H. Yang, *Blind Signal Separation and Extraction*, chapter 3 in "Unsupervised Adaptive Filtering" Volume I, edited by S. Haykin, Wiley, 2000.

[7] S. Cruces, A. Cichocki, and L. Castedo, "Blind source extraction in Gaussian noise," in *proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'2000), Helsinki, Finland*, June 2000, pp. 63–68.

[8] Xi-Ren Cao and Ruey wen Liu, "General approach to blind source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 562–571, Mar. 1996.

[9] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 3, no. 36, pp. 287–314, 1994.

[10] J. F. Cardoso, "Multidimensional independent component analysis," in *ICASSP*, 1998, vol. 4, pp. 1941–1944.

[11] H. H. Yang and S. Amari, "Adaptive on-line learning algorithms for blind source separation – maximum entropy and minimum mutual information," *Neural Computation*, 1997.

[12] D. Donoho, *On Minimum Entropy Deconvolution*, Applied Time Series Analysis II, D. F. Findley Editor, Academic Press, New York, 1981.

[13] J. H. Friedman, "Exploratory projection pursuit," *American Statistics Association*, vol. 82, no. 397, pp. 249–266, March 1987.

[14] R.A. Wiggins, *Minimum Entropy Deconvolution*, Geoexploration, 16, 1978.

[15] S. Amari, "Natural gradient learning for over- and under-complete bases in ica," *Neural Computation*, vol. 11, pp. 1875–1883, 1999.

[16] S. Cruces, *Una visión unificada de los algoritmos de separacion ciega de fuentes, An unified view of blind source separation algorithms*, Ph.D. thesis, University of Vigo. Signal Processing Dept., Spain, 1999.

[17] J.F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, dec 1996.

[18] S. C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Transactions on Neural Networks*, vol. 11, no. 6, pp. 1490–1497, November 2000.

[19] A. Schroeder J.H. Friedman, W. Stuetzle, "Projection pursuit density estimation," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 599–608, September 1984.

[20] S. Cruces, A. Cichocki, and S i. Amari, *The Minimum Entropy and Cumulant Based Contrast Functions for Blind Source Extraction*, pp. 786–793, Lecture Notes In Computer Science 2085, J. Mira & A. Prieto editors, Springer, 2001.

[21] S. Cruces, A. Cichocki, and L. Castedo, "An iterative inversion approach to blind source separation," *IEEE Transactions on Neural Networks*, vol. 11, no. 6, pp. 1423–1437, Nov. 2000.