4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

TWO-PHASE NONPARAMETRIC ICA ALGORITHM FOR BLIND SEPARATION OF INSTANTANEOUS LINEAR MIXTURES

*Chin-Jen Ku   Terrence L. Fine**

School of Electrical and Computer Engineering
Cornell University. Ithaca, New York 14853 U.S.A.
e-mails: kucj@ece.cornell.edu, tlfine@ece.cornell.edu

ABSTRACT

We propose a nonparametric independent component analysis (ICA) algorithm for the problem of blind source separation with instantaneous, time-invariant and linear mixtures. Our Init-NLE algorithm combines minimization of correlation among nonlinear expansions of the output signals with a good initialization derived from search guided by statistical tests for independence based on the power-divergent family of test statistics. Such initialization is critical to reliable separation. The simulation results obtained from both synthetic and real-life data show that our method yields consistent results and compares favorably to the existing ICA algorithms.

1. INTRODUCTION

The problem of blind source separation (BSS) arises when one desires to extract certain signals based solely on their mixture(s). We present a new two-phase nonparametric BSS algorithm enabling reliable recovery of the original signals by iteratively minimizing the correlation among nonlinear expansions of the output signals starting from a carefully chosen initial condition. An important element of our success is the implementation of an independence test based on the so-called *power-divergent (PD) family* of test statistics [12] which helps us find a suitable initial starting point. This paper is organized as follows. In section 2 we review the BSS problem formulation and the current state-of-the-art algorithms. In sections 3,4 we derive the iteration formula performing decorrelation of the nonlinear expansions of a set of signals and show how to implement an independence test using the PD test statistics. In section 5 we explain how to construct the two-phase ICA algorithm by implementing the idea from section 4 to find a suitable initial starting point (phase *Init*) for the decorrelation algorithm (phase *NLE*) described in section 3. In section 6, we define a new performance index for source separation and show that another standard performance index, *Amari error* [5], can be insensitive to demixing failures. In section 7 we compare the performance of our method to several existing ICA algorithms using both synthetic and benchmark data sets.

* Corresponding authors, research of both authors partially supported by AFOSR Grant F30602-00-2-0558

2. BSS: BACKGROUND

2.1. Mathematical Formulation

Our study is focused on the problem of source separation for *instantaneous* and *time-invariant linear* mixtures. Define N as the number of source signals, M as the number of observed signals and L as the number of samples. One may express the linear mixing process as

$$X = A \cdot S \tag{1}$$

where S is an $N \times L$ matrix representing samples of N independent sources, A is an $M \times N$ real constant matrix with rank N ($M \geq N$), usually called the *mixing matrix*, and X is an $M \times L$ matrix of observed mixtures. Without loss of generality, we assume that the sources are zero-mean. The demixing process recovers the N independent sources (up to scaling and permutation) as Y by multiplying the observed mixtures X by an $N \times M$ demixing matrix W:

$$Y = WX = WAS = CS \tag{2}$$

We assume absence of noise and present the case of equal number of sources and received signals (N=M) that is known at the processing unit. The BSS problem can be translated into *Independent Component Analysis* (ICA) for the ultimate goal consists of recovering statistically independent outputs.

2.2. Prior work on ICA

Recently ICA has received considerable attention that has resulted in the creation of several successful ICA algorithms. We distinguish in particular the *InfoMax* algorithm [6] which attempts to maximize the information transfer from input to output using a neural network framework, the JADE algorithm [1] performing joint approximate diagonalization of the eigen-matrices of the cumulant tensor [10], and the fixed-point ICA algorithm (Fast-ICA) [3] which tries to minimize some contrast function approximating the mutual information among the signals. Fast-ICA algorithm is currently a state-of-the-art algorithm in the field.

3. NLE ALGORITHM

3.1. Objective Function

It is well known from probability theory that two scalar random variables X_1, X_2 are statistically independent if and only if $f(X_1)$ and $g(X_2)$ remain uncorrelated for all f and g ranging over a separating class of functions [9, 11]. Hence, we may define a measure of dependence based on the correlation among $f(X_1)$ and $g(X_2)$ with adequate selection of the functions f, g. Given a demixing matrix W and the corresponding output signals Y, our new algorithm first expands Y to $Z_1, Z_2, \ldots Z_F$ where $Z_i = f_i(Y)$ such that $f_1(Y) = Y$ and f_i being some nonlinear functions applied elementwise to Y for $i \geq 2$. This results in the following $NF \times L$ compound matrix

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_F \end{pmatrix} = \begin{pmatrix} f_1(Y) \\ \vdots \\ f_F(Y) \end{pmatrix} \tag{3}$$

Its covariance matrix is given in block form by

$$C_{\mathbf{Z}} = \begin{pmatrix} Cov(Z_1, Z_1) & \ldots & Cov(Z_1, Z_F) \\ \vdots & & \vdots \\ Cov(Z_F, Z_1) & \ldots & Cov(Z_F, Z_F) \end{pmatrix}, \tag{4}$$

where the covariance matrices $\{Cov(Z_i, Z_j)\}$ are approximated by time averaging over the data samples. In order to take into account the cross-covariances only with respect to pairs of distinct variables $\{(y_i, y_j) : i \neq j\}$, we first form the symmetric matrix $\mathcal{M}$ defined as

$$\mathcal{M} = M \otimes C_{\mathbf{Z}} \tag{5}$$

The operator $\otimes$ denotes elementwise matrix multiplication and $M = \mathbf{1} - \mathbf{I}$, where $\mathbf{1}$ is an $NF \times NF$ matrix with all entries equal to one and $\mathbf{I}$ is an $NF \times NF$ matrix comprised of F^2 $N \times N$ identity matrices:

$$M = \mathbf{1} - \begin{pmatrix} I_N & \ldots & I_N \\ \vdots & & \vdots \\ I_N & \ldots & I_N \end{pmatrix} \tag{6}$$

We select as our objective function the sum of squared covariances of all pairs of variables $(f_p(y_i), f_q(y_j))$ for $i \neq j$ and express it as

$$\mathcal{L} = \frac{1}{2} \cdot trace(\mathcal{M} \cdot \mathcal{M}) \tag{7}$$

The estimation of the demixing matrix W through decorrelation of nonlinear expansions is approached in a more general fashion by Bach and Jordan [2], although their Kernel-ICA method carries a large computational burden.

3.2. First derivative of cost function

From (7), one can derive the gradient of $\mathcal{L}$ with respect to W as follow. First let $\mathcal{K}_{ij}$ denote $\nabla_W(\mathcal{M}_{ij})$, so that

$$\nabla_W \mathcal{L} = \sum_{i,j=1}^{NF} \mathcal{M}_{ij} \cdot \mathcal{K}_{ij} \tag{8}$$

According to (5), if we first rewrite the indices i,j as

$$\begin{aligned} i &= p + (l-1)N \\ j &= q + (m-1)N \end{aligned} \tag{9}$$

with $1 \leq l, m \leq F$ and $1 \leq p, q \leq N$, $\mathcal{M}_{ij}$ can be explicitly expressed as

$$\mathcal{M}_{ij} = \begin{cases} 0 & \text{if } p = q \\ Cov(f_l(y_p), f_m(y_q)) & \text{if } p \neq q \end{cases} \tag{10}$$

Let $\overline{f_l}(y_p)$, $\overline{f_m}(y_q)$ denote the sample averages of $f_l(y_p(t))$ and $f_m(y_q(t))$, respectively. We obtain for $p \neq q$

$$\begin{aligned} \mathcal{K}_{ij} =& \nabla_W Cov(f_l(y_p), f_m(y_q)) \\ =& \nabla_W [\frac{1}{L} \sum_{t=1}^{L} (f_l(y_p(t) - \overline{f_l}(y_p)) \cdot \\ & (f_m(y_q(t) - \overline{f_m}(y_q))] \end{aligned} \tag{11}$$

The gradient of $f_l(y_p(t))$ with respect to W is equal to

$$\begin{aligned} \nabla_W f_l(y_p(t)) &= f_l'(y_p(t)) \cdot \nabla_W y_p(t) \\ &= f_l'(y_p(t)) \underline{\mathbf{e}}_p X(\cdot, t)^T \end{aligned} \tag{12}$$

where $\underline{\mathbf{e}}_i$ is the i^{th} column of identity matrix I_N. Using the following notation for the $NF \times L$ matrices

$$D = \begin{pmatrix} f_1'(Y) \\ \vdots \\ f_F'(Y) \end{pmatrix} \tag{13}$$

$$\tilde{Z} = Z - \overline{Z} = \begin{pmatrix} f_1(y_1) - \overline{f_1}(y_1) \\ \vdots \\ f_F(y_N) - \overline{f_F}(y_N) \end{pmatrix} \tag{14}$$

representing the derivatives and zero-mean versions of non-linear function expansions, it follows that

$$\mathcal{K}_{ij} = [\underline{\mathbf{e}}_p(D(i,\cdot)\otimes\tilde{Z}(j,\cdot)) + \underline{\mathbf{e}}_q(\tilde{Z}(i,\cdot)\otimes D(j,\cdot))]\cdot X^T/L \qquad (15)$$

Thus the demixing matrix W may be estimated iteratively using standard gradient-based learning rules, in particular steepest descent algorithm:

$$\hat{W}(k+1) = \hat{W}(k) - \eta(k)\cdot\nabla_W\mathcal{L}(k) \qquad (16)$$

In the context of independent source separation, however, a successful result requires a good choice of initial condition for the iteration (16). We now turn to a method to provide such initial conditions.

4. TESTS FOR STATISTICAL INDEPENDENCE

4.1. Power-divergence family test statistics

In discrete multivariate analysis, it is known that a hypothesis regarding the statistical properties among N discrete variables $y_1\dots y_N$ can be tested by evaluating the chi-square goodness-of-fit statistics on the corresponding contingency table. Assume that the i^{th} variable y_i has been quantized into q intervals or cells $\{J_j^{(i)} : j = 1\dots q\}$ for $i = 1\dots N$. We define the *cell counting variables*

$$N_{\mathbf{j}} = N(\{j_1, j_2 \dots j_N\}) \qquad (17)$$

as the number of data points where we have concurrently $y_1 \in J_{j_1}^{(1)}, \dots, y_N \in J_{j_N}^{(N)}$, with the argument $\mathbf{j}$ taking values in $\mathcal{J} = \{1\dots q\}^N$. Read and Cressie [12] introduced the *power-divergence (PD) family* of goodness-of-fit statistics to measure the deviation of observed data from the stipulated hypothesis:

$$2I^\lambda(\mathbf{X}:\mathbf{m}) = \frac{2}{\lambda(\lambda+1)}\sum_{\mathbf{j}\in\mathcal{J}} N_{\mathbf{j}}\cdot[(\frac{N_{\mathbf{j}}}{m_{\mathbf{j}}})^\lambda - 1] \qquad (18)$$

where $m_{\mathbf{j}}$ is the expected number of data belonging to the cell $\mathbf{j}$ under the hypothesis and $N_{\mathbf{j}}$ is the number of data actually observed. When the expected cell counts $\{m_{\mathbf{j}}\}$ are unknown they are replaced by their estimates $\{\hat{m}_{\mathbf{j}}\}$ based on the observed data. Under the assumptions of independence among the data points and fixed cell probabilities for all L, all the PD family statistics are asymptotically equivalent when the hypothesis is true. The common asymptotic distribution is χ^2 with a number of degrees of freedom depending on the total number of cells and estimated parameters. Numerous researchers suggest that choosing λ between 0 and 1 [12](pp. 63) (*e.g.* $\lambda = \frac{2}{3}$) has the best detection power against arbitrary lack of fit to the hypothesis.

4.2. Adaptation to ICA problems

We use the PD test statistics described above to determine whether a given demixing matrix recovers statistically independent signals even when the individual signal samples $\{s_i(t)\}$ are not *i.i.d.* If the signals $\{s_i(t)\}$ are stationary and mutually independent of each other, the marginal sums of cell counting variables $\{N_{i+}, N_{+j}\}$ determine the unbiased estimates of expected cell counts $\mathbf{m}$. For $N = 2$, we have

$$\hat{m}_{ij} = \frac{N_{i+}\cdot N_{+j}}{L},\ \ (i,j)\in\{1,2\dots q\}^2 \qquad (19)$$

with $N_{i+} = \sum_{j=1}^q N_{ij}$ and $N_{+j} = \sum_{i=1}^q N_{ij}$. When the individual signal samples are i.i.d, $\{N(\mathbf{j})\}$ are distributed according to a multinomial distribution and the estimates in (19) become the maximum likelihood estimates of $\mathbf{m}$.

Extension of equation (19) to higher dimensional cases is straightforward. Moreover, it is advised by many to define *equiprobable* cells to guarantee the unbiasness of PD test statistics [12]. In this case, the estimates of expected cell counts are equal to the ratio between sample size and the total number of cells:

$$\hat{m}_{\mathbf{j}} = \frac{L}{q^N}\quad \forall\mathbf{j}\in\mathcal{J} \qquad (20)$$

The independence test can be easily performed by substituting the value of $\hat{\mathbf{m}}$ from (20) into (18). In the case of *i.i.d* samples, the number of degrees of freedom of the approximating chi-square distribution is equal to $(q^N - qN + N - 1)$ by taking into account all the constraints on the marginal sums. Therefore, a significance level may be established to verify the independence hypothesis.

5. INITIALIZATION

Preliminary simulation results show that our algorithm in section 3 sometimes encounters convergence difficulty or spurious solutions when the initial guess is far away from a demixing solution. To improve its robustness, we propose a two-phase ICA algorithm, called *Init-NLE*. The first phase *Init* uses a power divergence statistics-based independence test to obtain a suitable initialization of demixing matrix W_0, whereupon the second phase *NLE* uses (16) to reach a demixing matrix with high separation performance. Search for W_0 may be carried out using an orthogonal ICA approach [4, 7] which is based on the eigen-decomposition of the covariance matrix of the observed mixtures:

$$R_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^{\mathbf{T}}\} = AR_{\mathbf{s}}A^T = ELE^T \qquad (21)$$

where $R_{\mathbf{s}}$ is diagonal under hypothesis of statistical independence among the sources, E is an orthogonal matrix containing the eigenvectors of R_X, and L is a diagonal matrix with all positive entries on the diagonal since A is assumed invertible and R_X positive definite. From (21), one can show that there exists an orthogonal matrix U satisfying

$$A = EL^{\frac{1}{2}} U R_{\mathbf{s}}^{-\frac{1}{2}} \tag{22}$$

In virtue of the scaling and permutation ambiguity in blind identification [4], the demixing matrix may be estimated by

$$\hat{W} = U^T L^{-\frac{1}{2}} E^T \tag{23}$$

The matrix $Y = \hat{W}X$ recovers S to within a rescaling of rows by $R_{\mathbf{s}}^{-\frac{1}{2}}$ and a permutation and sign change determined by the choice of U. Hence, the initialization problem is reduced to finding an orthogonal transformation U such that the resulting output Y yields low PD test statistics. The corresponding demixing matrix is then used to initialize the second phase.

6. PERFORMANCE EVALUATION

We introduce an index measuring the source separation quality based on the *signal-to-interference energy ratio* (SIR). Let C be the global transfer matrix between original and estimated sources as defined in (2), Es_i^2 be the energy of i^{th} original source and $E_{i,j}$ the energy of i^{th} source recovered on j^{th} channel. The quantity

$$\gamma_{j,i} = \frac{E_{i,j}}{\sum_k E_{k,j}} = \frac{C_{ji}^2 Es_i^2}{\sum_k C_{jk}^2 Es_k^2} \tag{24}$$

represents the ratio between the energy of i^{th} source recovered on channel j and the total energy recovered on channel j. Now let

$$\gamma_i = \max_j \gamma_{j,i}, \quad SIR_i(dB) = 10 \log_{10} \frac{\gamma_i}{1-\gamma_i} \tag{25}$$

then SIR_i is the ultimate quality of recovery with respect to the i^{th} source. If SIR_i is positive, the i^{th} signal is dominant at some output channel of the demixer. Conversely, if SIR_i is negative, the i^{th} signal does not dominate at any of the output channels. Such a case happens when the recovery is *incomplete*. We will conservatively use the minimum SIR over all the channels (MSIR) as the global separation performance index. The performance index MSIR tends to infinity when C belongs to the family of ideal separating matrices $\mathcal{I}$ (product of scaling and permutation matrices).

It is worth mentioning another standard performance index, *Amari error* [5], that is widely used in the ICA community. This index is nonnegative and reaches zero when C is an ideal separating matrix. However, Amari error may be misleading in that it doesn't always reflect the true quality of source separation. Consider the following global transfer matrix:

$$C_1 = \begin{pmatrix} \epsilon & 1 \\ \epsilon^2 & \epsilon \end{pmatrix}, \quad 0 < \epsilon \ll 1 \tag{26}$$

Although associated with a small Amari error (e.g. 0.05 when $\epsilon = 0.05$), C_1 is far away from the family $\mathcal{I}$. It recovers with great precision the second component (twice) but completely discards the first source signal. This is an instance of *incomplete recovery* which is undesirable for ICA algorithms. Inconsistency may also occur when the source signals have disparate power magnitude. Let two sources $S = \{s_1, s_2\}$ have average energy 1 and 10 and consider respectively the cases

$$C_2 = \begin{pmatrix} 1 & 0.1 \\ 0 & 1 \end{pmatrix}, \quad C_3 = \begin{pmatrix} 1 & 0.01 \\ 0.3 & 0.7 \end{pmatrix} \tag{27}$$

Even though the corresponding Amari errors are 0.05 for C_2 and 0.1882 for C_3, the recovered sources in the second case are much closer to the original ones than in the first case. Using MSIR criterion, we obtain MSIR(C_2,S)=10dB and MSIR(C_3,S)=17.43dB, which is consistent with the quality of separation.

7. EXPERIMENTAL RESULTS

7.1. Simulation setup

We illustrate the performance of our new algorithm Init-NLE on several data sets. The first data set consists of one *i.i.d.* Rayleigh(1) distributed random sequence and one *i.i.d.* Laplacian(1) distributed sequence. The four other data sets are taken from ICALAB benchmark sets [8]:

- Data set 2: *Speech4.mat* (sources 1,2)
- Data set 3: *Sergio7.mat* (sources 3,6)
- Data set 4: *10halo.mat* (sources 1,3)
- Data set 5: *Gnband.mat* (sources 1,4)

Given the large size of data set 3 (L=10^4), we select the first 2000 samples from each source as the original signals. As for data set 4, which consists of 2 different human voices pronouncing the same word "hello", we use the first 3000 time samples from each source as they cover exactly the relevant part of the original data. It remains to choose the nonlinear functions in the NLE algorithm. These functions should be defined such that no one can be closely approximated by a linear combination of the others. Just as an arbitrary choice, the following nonlinear functions are used to experiment with our proposed Init-NLE algorithm:

$$f_2(x) = \frac{1}{1+e^{-x}}, f_3(x) = \cos 3x, f_4(x) = \sin 2x \tag{28}$$

The performance of Init-NLE is compared to three other standard ICA algorithms: InfoMax, Fast-ICA and JADE [1, 3, 6]. Fast-ICA algorithms using three different nonlinear functions are considered: cubic (*pow3*), square (*skew*) and $\tanh$. All the experiments conducted on the data sets are repeated over 200 randomly selected mixing matrices A and the average MSIR (dB) over these choices are reported.

7.2. Separation Performance

In Figure 1 we first illustrate the behavior of the PD test statistics in terms of the quality of demixing matrix W for a simple two *i.i.d* sources BSS problem. Four quantization levels ($q = 4$) are used for each signal to generate the contingency table. The values of PD test statistics and MSIR (dB) are plotted against the rotation angle θ defining U. We also plotted the 1%-significance level in the plot of PD test statistics vs. θ. We see readily that as the degree of separation improves, the PD test statistic decreases and stays below significance level when the separation is maximal. Search for an initial rotation matrix can be carried out by simple grid search for $N = 2, 3$ until the test statistic falls below significance, for instance. Another type of search is needed for larger N for computational efficiency. Figure 2 shows the performance realized by each of the four ICA algorithms on the first data set. Our algorithm clearly demonstrates better performance. In addition, we noted that InfoMax and Fast-ICA algorithms yielded unsatisfactory results for several of our randomly chosen mixing matrices as the resulting MSIR was below 5dB.

The data sets 2 to 5 deal with non *i.i.d* data samples. Unlike the previous case, we do not have enough information to determine the PD test statistic significance level and cannot use it as a stopping criterion in the search for an adequate orthogonal matrix. Nonetheless, we observe from many empirical simulation results based on nonstationary signals that the relationship between the test statistics and MSIR is maintained in that small values of the PD test statistics still correspond in most cases to high values of MSIR. Therefore, the initialization consists of search in the space of orthogonal matrices for the one with the smallest test statistic.

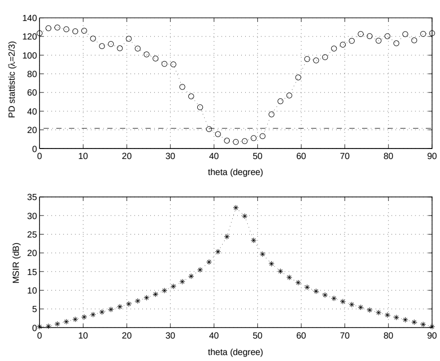

Fig. 1. Behavior of PD test statistic vs. quality of demixing matrix W. PD statistic is below significance level at maximal degree of separation (highest MSIR).

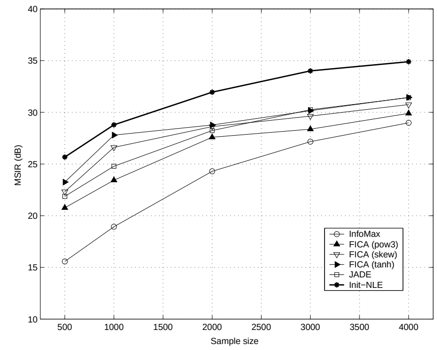

Fig. 2. Data set 1: plot of average MSIR versus sample size L over 200 randomly selected mixing matrices A.

Table 1. Data set 2: average MSIR (dB) vs. sample size L.

L	InfoMax	FICA (pow3)	FICA (skew)
500	20.66	31.26	25.12
1000	30.29	32.53	30.99
2000	19.44	24.09	25.44
3000	29.38	32.44	22.01
4000	31.46	**34.01**	22.99
5000	30.97	31.92	24.62
	FICA (tanh)	JADE	Init-NLE
500	**37.51**	31.37	34.85
1000	32.33	35.08	**36.52**
2000	25.21	21.19	**28.49**
3000	34.47	26.23	**38.81**
4000	33.68	30.58	33.38
5000	33.17	31.58	**33.47**

The corresponding experimental results are reported in Tables 1 to 4. We note in particular the performance realized by Init-NLE on data set 5, which contains two fourth order colored sources with a distribution close to Gaussian and is known to be a “hard” BSS benchmark. Init-NLE achieves much better separation than Fast-ICA and JADE. Overall, our algorithm succeeds in finding consistent demixing solution with superior performance compared to the three other ICA algorithms.

8. CONCLUSION

We have proposed a powerful nonparametric two-phase Init-NLE algorithm in the context of blind source separation with instantaneous, time-invariant and linear mixtures. It first deploys a statistical independence test to find an initial estimate of demixing matrix achieving sufficient degree of independence among the recovered signals. Subsequently, it refines the solution by minimizing the correlation among

Table 2. Data set 3: average MSIR (dB) vs. sample size L.

L	InfoMax	FICA (pow3)	FICA (skew)
500	28.11	36.49	36.65
1000	33.26	36.28	34.47
1500	34.54	29.71	30.20
2000	35.05	31.88	30.46
	FICA (tanh)	JADE	Init-NLE
500	34.81	39.04	**44.83**
1000	35.14	40.36	**50.15**
1500	35.41	31.65	**39.22**
2000	31.20	35.22	**38.51**

Table 3. Data set 4: average MSIR (dB) vs. sample size L.

L	3000	2000	1000	500
InfoMax	15.76	19.35	16.23	10.60
FICA (pow3)	20.89	26.74	19.76	1.27
FICA (skew)	18.10	19.34	**31.42**	0.39
FICA (tanh)	15.54	25.59	17.20	6.24
JADE	24.06	**35.85**	18.67	0.73
Init-NLE	**29.83**	33.50	19.77	**31.67**

different nonlinear expansions of all recovered signals. The experimental results show that our method provides consistent results for a variety of source distributions and can outperform the existing state-of-the-art ICA algorithms in some difficult problems. Furthermore, the proposed PD independence test may be regarded as a tool to evaluate the quality of solution provided by any ICA algorithm without knowing the mixing matrix or the sources signals. Our recent efforts have been focused on its extension to multiple sources problems where time-efficient search for a good initialization is essential. At this point we have been able to perform BSS for mixtures of up to 5 sources. Further investigations including efficient numerical implementation of the algorithm for large-scale problems and determination of the significance level for non *i.i.d.* signals may help our new algorithm achieve faster reliable source separation.

9. REFERENCES

[1] Jean-François Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, September 2001.

[2] Francis R. Bach and Micheal I. Jordan, "Kernel Independent Component Analysis," *Journal of Machine Learning Research*, 3, pp. 1-48, 2002.

[3] Aapo Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.

Table 4. Performance results for data set 5 (L=2000).

InfoMax	FICA (pow3)	FICA (skew)
13.12	5.32	14.21
FICA (tanh)	JADE	Init-NLE
9.98	5.43	**18.56**

[4] Jean-François Cardoso, "Blind signal separation: Statistical Principles," *Proceedings of the IEEE*, vol. 86, no. 10, October 1998.

[5] S. Amari, A. Cichocki and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems 8*, MIT Press, 1996.

[6] A. Bell, T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1004-1034, 1995.

[7] Jean-François Cardoso, "On the performance of orthogonal source separation algorithms," *EUSIPCO*, (Edinburgh), pp. 776-779, September 1994.

[8] A. Cichocki, S. Amari, K. Siwek et al., "ICALAB for Signal Processing - benchmarks", *http://www.bsp.brain.riken.go.jp/ICALAB*.

[9] L. Breiman, *Probability*, Classics in Applied Mathematics. SIAM, 1992.

[10] Jean-François Cardoso and Antoine Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362-370, December 1993.

[11] A. Feuerverger, "A consistent test for bivariate dependence," *International Statistical Review*, vol. 61, no. 3, pp. 419-433, 1993.

[12] Timothy R. C. Read and Noel A. C. Cressie, *Goodness-of-fit statistics for discrete multivariate analysis*, Springer-Verlag, 1988