NON-PARAMETRIC ICA

Riccardo Boscolo[*]*, Hong Pan*[‡]*, and Vwani P. Roychowdhury*[*]

[*] Electrical Engineering Department
University of California, Los Angeles
Los Angeles, CA 90095
{riccardo,vwani}@ee.ucla.edu

[‡] Functional Neuroimaging Laboratory
Department of Psychiatry
Weill Medical College of Cornell University
New York, NY 10021
hop2001@med.cornell.edu

ABSTRACT

We introduce a novel approach to the blind signal separation (BSS) problem that is capable of *jointly* estimating the probability density function (pdf) of the source signals and the unmixing matrix. We demonstrate that, using a kernel density estimation based Projection Pursuit (PP) algorithm, it is possible to extract, from instantaneous mixtures, independent sources that are arbitrarily distributed. The proposed algorithm is non-parametric, and unlike conventional Independent Component Analysis (ICA) frameworks, it requires neither the definition of a contrast function, nor the minimization of the high-order cross-cumulants of the reconstructed signals. We derive a new method for solving the resulting constrained optimization problem that is capable of accurately and efficiently estimating the unmixing matrix, and which does not require the selection of any tuning parameters. Our simulations demonstrate that the proposed method can accurately separate sources with arbitrary marginal pdfs with significant performance gain when compared to existing ICA algorithms. In particular, we are successful in separating mixtures of skewed, almost zero-kurtotic signals, which other ICA algorithms fail to separate.

1. INTRODUCTION

Regardless of the nature of the unmixing matrix estimation technique (maximum likelihood, stochastic gradient, or batch mode optimization), the problem of choosing a suitable model for the pdf of the sources to be reconstructed is pivotal in most BSS methods. Whether the objective function for these algorithms is derived from the InfoMax principle [1] or from the Redundancy Reduction principle [2], a method to estimate the differential entropy of the projected data has to be devised. Several non-linear contrast functions have been proposed, each capable of modeling a specific class of source signals (see [3] for a survey of such functions). Moreover, a technique that adapts the sign of the contrast function according to the estimated fourth-order moments of the sources has been proposed [4].

As an alternative, BSS frameworks that are uniquely based on the simultaneous diagonalization of the higher-order cumulant tensors of the projected data have been developed [5]. These methods are not restricted by a specific choice for the parametrization of the pdf of the source signals. Potential limitations affecting algorithms based on this approach include their sensitivity to outliers and their reliance primarily on fourth-order cumulants, to perform the separation.

Recently, new methods that employ more flexible adaptive models for the pdf of the source signals have been proposed [6, 7]. Although these frameworks are less stiff in defining a contrast function, they are still not capable of universally modeling arbitrarily distributed sources.

In this paper we show that using a Projection Pursuit [8] framework based on the order-1 entropy index and a robust kernel density estimation technique, it is possible to derive a novel blind source separation algorithm, which is capable of reliably and accurately separating sources with arbitrary marginal distributions. The resulting algorithm is non-parametric, data-driven, and *does not require the definition of a contrast function*.

Projection Pursuit (PP) is a statistical exploratory technique, whose goal is the identification of 'interesting' low-dimensional linear projections of high-dimensional data, based on certain (usually non-linear) index functional [9, 10]. Although certain connections have been recognized [11] between the Projection Pursuit algorithms based on an index that measures deviation from gaussianity ("negentropy index") and the problem of blind signal separation, these results have been exploited mainly to justify common choices for the contrast function used in several ICA algorithms [12]. No true PP based blind signal separation algorithm has been proposed and successfully implemented.

In Section 2 we establish a PP framework for blind signal separation. The idea that is pivotal to the proposed method is described in detail in Section 3. Section 4 pre-

sents the results of our simulations. Final remarks and conclusions are given in Section 5.

2. PROJECTION PURSUIT AND BLIND SIGNAL SEPARATION

We make the conventional assumption that N independent sources $\mathbf{s} = \{s_1, \ldots, s_N\}$, are instantaneously mixed by a full rank mixing matrix A, giving $\mathbf{x} = A\mathbf{s}$. The recovery of these sources is attempted by linear projection through an unmixing matrix W, in such a way that the random variables $\mathbf{y} = W\mathbf{x}$, in the case of perfect separation, represent a scaled and permuted version of the original sources. Assuming that M samples of $\mathbf{x}$ are available as a matrix X, it is convenient to sphere and center the data by principal component analysis. If the recovery of the original sources is attempted using the sphered data matrix, and if the signals are truly uncorrelated, then it can be shown that the unmixing matrix must be orthogonal (a simple proof of this result is given in [13]). Throughout this paper we will make the assumption that the data is sphered, and that the problem is reduced to the estimation of an orthogonal matrix W, although this is not strictly required in our algorithm.

A method to seek a single one-dimensional projection of N-dimensional data was developed by Jones in [8]. A Projection Pursuit algorithm based on the order-1 entropy index of interestingness, identifies such a projection solving the following problem:

$$\max_{\mathbf{w}} \; -H(y) = \max_{\mathbf{w}} \int p(y) \log p(y) dy \qquad (1)$$
$$\text{s.t. } ||\mathbf{w}|| = 1.$$

The constraint $||\mathbf{w}|| = 1$ restricts the search space to linear projections that preserve the variance of the original data. This restriction also ensures that $H(y)$ is bounded and that the optimization is well posed. We can define an extension of problem (1) to seek N linearly independent projection vectors, as follows:

$$\max_{W} \; -\sum_{i=1}^{N} H(y_i) \qquad (2)$$
$$\text{s.t. } \det W \neq 0$$
$$||\mathbf{w}_i|| = 1 \; , \quad i = 1, \ldots, N \; ,$$

where $\mathbf{w}_i$ are the rows of the matrix W and $y_i = \mathbf{w}_i \mathbf{x}$. The constraint $\det W \neq 0$ guarantees that the N projection directions are linearly independent. As an alternative, the projection directions could be identified one at the time, using a structure removal approach [8] that prevents the algorithm from choosing the same projection direction twice.

Instead of dealing with this strict constraint, we can solve the relaxation of problem (2), defined by:

$$\max_{W} \; -\sum_{i=1}^{N} H(y_i) + \log|\det W| \qquad (3)$$
$$\text{s.t. } ||\mathbf{w}_i|| = 1 \; , \quad i = 1, \ldots, N \; . \qquad (4)$$

The term $\log|\det W|$ guarantees that the matrix W is full rank for any feasible solution to the problem. Moreover, because of the constraints $||\mathbf{w}_i|| = 1$, combined with the Hadamard inequality:

$$|\det W| \leq \prod_{i=1}^{N} ||\mathbf{w}_i|| = 1 \; , \qquad (5)$$

we have that:

$$\log|\det W| \leq 0 \; , \qquad (6)$$

where equality holds when W is orthonormal. Therefore, the two problems (2) and (3) are asymptotically equivalent when sphered measurement data is used to estimate the unmixing matrix, and the sources are truly uncorrelated. In this case, in fact, the term $\log|\det W|$ in (3) forces the iterates of the solution to remain inside the feasible set, and it approaches zero when the optimal solution is attained.

We can show that the modified Projection Pursuit framework defined by (3), satisfies the basic assumption behind all ICA methods that the statistical mutual dependence between the reconstructed signals y_i is minimized. If we add to the objective function in (3) a constant term equal to the entropy of the measurement data $H(\mathbf{x})$, we can rewrite it as:

$$\max_{W} \; -\sum_{i=1}^{N} H(y_i) + \log|\det W| + H(\mathbf{x}) \qquad (7)$$
$$= \min_{W} \; I(y_1, \ldots, y_N) \; ,$$

since $H(\mathbf{y}) = H(\mathbf{x}) + \log|\det W|$, thus proving that the optimization problem (3) satisfies the redundancy reduction principle [2].

Although the framework defined by the objective function (3) bears similarities with other ICA methods, the proposed algorithm presents several novel aspects:

- The pdf of the reconstructed source signals p_{y_i} is *estimated directly from the data*, rather than being modeled using a fixed or adaptive parametrization. As a consequence, the estimation of the unmixing matrix does not require the definition of a contrast function.
- The constraints (4) guarantee that the method is a *true* Projection Pursuit algorithm, which seeks the set of N projections maximizing the total negative entropy, while preserving the data variance .

- A new algorithm that is capable of efficiently enforcing the constraints is developed. The resulting optimization technique *does not require the selection of any learning parameters*.

The next section is dedicated to the description of such aspects.

3. JOINT ESTIMATION OF UNMIXING MATRIX AND PDF OF THE SOURCE SIGNALS

The problem defined in (3) involves the constrained maximization of the following objective function:

$$\begin{aligned} L(W) &= -\sum_{i=1}^{N} H(y_i) + \log|\det W| \qquad (8) \\ &= \sum_{i=1}^{N} E\left[\log p_{y_i}(\mathbf{w}_i\mathbf{x})\right] + \log|\det W| \,, \end{aligned}$$

where $\mathbf{w}_i$ are the *rows* of the matrix W. Given a batch of data of size M, we can approximate the expectation in (8) with its ergodic average:

$$L(W) \approx \frac{1}{M}\sum_{i=1}^{N}\sum_{k=1}^{M} \log p_{y_i}(\mathbf{w}_i\mathbf{x}^{(k)}) + \log|\det W| \,, \quad (9)$$

where $\mathbf{x}^{(k)}$ is the kth column of the sphered data matrix. A universal model is proposed where the p_{y_i} are directly estimated from the data, using a kernel density estimation technique [14]. When a suitable kernel is chosen (see [14]) this estimator is asymptotically unbiased and efficient, and it is shown to converge to the true pdf under several measures. The marginal distribution of the source signals can be approximated as:

$$p_{y_i}(y_i) = \frac{1}{Mh}\sum_{m=1}^{M} \phi\left(\frac{y_i - Y_{im}}{h}\right) \,, \qquad (10)$$

where h is the kernel bandwidth and ϕ is a gaussian kernel:

$$\phi(u) \triangleq \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}} \,. \qquad (11)$$

The kernel centroids Y_{mi} are equal to:

$$Y_{im} = \mathbf{w}_i\mathbf{x}^{(m)} = \sum_{n=1}^{N} w_{in}x_{nm} \,. \qquad (12)$$

One of the advantages of the kernel estimate defined by (10) is that it is a continuous and differentiable function of the unmixing matrix elements w_{ij}. For example its gradient with respect to $\mathbf{w}_i$ can be written as:

$$\nabla_{\mathbf{w}_i} p(y_i) = \frac{1}{Mh^2}\sum_{m=1}^{M} \mathbf{x}_m (y_i - \mathbf{w}_i\mathbf{x}_m)\phi\left(\frac{y_i - \mathbf{w}_i\mathbf{x}_m}{h}\right) \,. \quad (13)$$

Evaluating the estimates of the pdf of the source signals at the data points, we obtain:

$$p_{y_i}(\mathbf{w}_i\mathbf{x}^{(k)}) = \frac{1}{Mh}\sum_{m=1}^{M} \phi\left(\frac{\mathbf{w}_i\left(\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\right)}{h}\right) \,. \quad (14)$$

If we write the objective (8) as follows:

$$L(W) = L_0(W) + \log|\det W| \,, \qquad (15)$$

then we can re-write L_0(W) replacing the marginal pdfs p_{y_i} with the kernel density estimates:

$$\begin{aligned} L_0(W) &= \sum_{i=1}^{N} E\ \log\left[\frac{1}{Mh}\sum_{m=1}^{M}\phi\left(\frac{y_i - Y_{im}}{h}\right)\right] \qquad (16) \\ &\approx \frac{1}{M}\sum_{i=1}^{N}\sum_{k=1}^{M}\log\left[\frac{1}{Mh}\sum_{m=1}^{M}\phi\left(\frac{\mathbf{w}_i\left(\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\right)}{h}\right)\right] \,. \end{aligned}$$

The optimization problem becomes:

$$\max_{W}\ \frac{1}{M}\sum_{i=1}^{N}\sum_{k=1}^{M}\log\left[\frac{1}{Mh}\sum_{m=1}^{M}\phi\left(\frac{\mathbf{w}_i\left(\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\right)}{h}\right)\right] + \log|\det W| \qquad (17)$$

$$\text{s.t.} \qquad ||\mathbf{w}_i|| = 1 \,, \quad i = 1, \ldots, N \,. \qquad (18)$$

Given the sample data $\mathbf{x}^{(k)}, k = 1, \ldots, M$, the objective (17) is a non-linear function only of the elements of the matrix W. The parameter h controls the smoothness of the functional and its optimal value is a function of the sample size ($h = 1.06M^{-1/5}$) [14]. Our simulation experiments show that variations up to $\pm 50\%$ from the default value do not affect the performance of the algorithm significantly. Using the FFT algorithm, the objective function is evaluated at a cost proportional to $\mathcal{O}(NM\log_2 M)$, while the N^2 derivatives can be computed with a number of operations proportional to $\mathcal{O}(N^2 M\log_2 M)$.

If a method is devised to remove the constraints (18), then the *unconstrained* optimization can be performed with a suitable algorithm, such as the Quasi-Newton method or the Conjugate Gradient algorithm. Such removal is achieved by operating the substitution:

$$\mathbf{w}_i = \frac{\tilde{\mathbf{w}}_i}{||\tilde{\mathbf{w}}_i||}, \quad i = 1, \ldots, N \,. \tag{19}$$

Using the transformation (19), the matrix W can be written as $W = \tilde{D}^{-1}\tilde{W}$, with:

$$\tilde{D} = \begin{bmatrix} ||\tilde{\mathbf{w}}_1|| & & 0 \\ & \ddots & \\ 0 & & ||\tilde{\mathbf{w}}_N|| \end{bmatrix}, \tag{20}$$

thus $\tilde{W} = \tilde{D}W$. Then:

$$\log|\det W| = -\sum_{i=1}^{N} \log||\tilde{\mathbf{w}}_i|| + \log|\det \tilde{W}| \,. \tag{21}$$

The derivatives with respect to $\tilde{w}_{ij}$ are thus computed as:

$$\frac{\partial(\log|\det W|)}{\partial \tilde{w}_{ij}} = -\frac{\tilde{w}_{ij}}{||\tilde{\mathbf{w}}_i||^2} + \left[(\tilde{W}^T)^{-1}\right]_{ij} \,. \tag{22}$$

When W is orthogonal ($W^{-1} = W^T$), we have:

$$(\tilde{W}^T)^{-1} = \tilde{D}^{-1}(W^T)^{-1} = \tilde{D}^{-2}\tilde{W} \,, \tag{23}$$

and the gradient coefficients in (22) are identically zero. Therefore, the second term of the cost function (17) no longer enters the optimization procedure if the matrix W is close to orthogonal. The calculation of the gradient of the first term of the cost function is slightly more involved and only the final result is reported in (24).

is capable of accurately estimating sources #3, #5 and #6, which the other algorithms fail to separate. The fact that distributions #3 and #6 are skewed and almost zero-kurtotic, results in a mismatch between their actual pdf and the parametric model assumed by the InfoMax algorithms. Moreover, the fact that only the fourth-order cumulants are used and the skewness information is ignored, might explain why Jade fails to separate these sources as well. Figure 2 shows the kernel density estimate of the pdf of source #3, as it is reconstructed by each algorithm, using 3000 data samples. Even for standard sub-gaussian or super-gaussian distributions (sources #1,#2, and #4), the proposed algorithm outperforms the existing methods, showing the benefits of accurately modeling the pdf of the source signals.

In a second experiment, we generated 100 realizations of each of five sources with skewness uniformly varying between 0 and 1, all with a theoretical kurtosis equal to 0.75. These synthetic signals were mixed, one at the time, with an independent gaussian source and the separation was attempted with each of the aforementioned algorithms. Ideally, an increase in the source skewness should facilitate the separation, because of the larger deviation from gaussianity. The results of this experiment (Table 2) demonstrate that standard ICA algorithms are not capable of accurately separating skewed sources, but, rather, their performance is inversely related to an increase in the skewness of the sources to be reconstructed. The proposed method, on the other hand, takes full advantage of the increased deviation from

$$\frac{\partial L_0(\tilde{W})}{\partial \tilde{w}_{ij}} = \frac{1}{M}\sum_{k=1}^{M} \frac{-\sum_{m=1}^{M}\left(X_{jk} - X_{jm} - \tilde{\mathbf{w}}_i(\mathbf{x}^{(k)} - \mathbf{x}^{(m)})\tilde{w}_{ij}\right)\tilde{\mathbf{w}}_i\left(\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\right)\phi\left(\frac{\tilde{\mathbf{w}}_i\left(\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\right)}{h}\right)}{h^2 \cdot \sum_{m=1}^{M}\phi\left(\frac{\tilde{\mathbf{w}}_i\left(\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\right)}{h}\right)} \tag{24}$$

4. SIMULATION RESULTS

In order to evaluate the proposed algorithm, we performed a first simulation experiment where 1000 independent realizations of the six pdfs listed in Table 1 were generated, with sample sizes ranging from 500 to 5000. These synthetic sources were mixed using randomly generated, full rank matrices (condition number ≤ 20), and the separation was attempted with each of the following algorithms: the original InfoMax ICA [1], the Extended InfoMax ICA [4], Cardoso's Jade [5], and our algorithm. The software for these ICA algorithms was downloaded from the webpages of the authors.

The results of this first experiment clearly show that the proposed algorithm outperforms the other ICA methods (Figure 1). In particular, only the non-parametric ICA gaussianity to improve the separation performance. Because it is not known a-priori whether the source signals have an asymmetric distribution or not, it is not possible in general to remove the skewness through pre-processing of the data. Therefore, only a non-parametric universal model for the pdf of the sources guarantees an accurate separation, when no a-priori knowledge on such pdfs is available.

In a third experiment, we evaluated the convergence properties of each ICA algorithm. The goal was to measure the approximate number of data samples required by each method to achieve a median SNR of at least 20dB. For this purpose, we created mixtures of four independent sources with super-gaussian (kurtosis ≈ 2.2) symmetric pdf and we averaged the separation results over 100 simulations, for different sample sizes. The choice of standard super-gaussian sources guarantees that the experiment is unbiased, since all ICA algorithms are capable of separating this type

Table 1. Distribution of the synthetic sources used in the first simulation experiment (see [15] for a description of the distributions generated with the Power Method).

Source#	Source type	Skewness	Kurtosis
1	Power Exponential Distribution (α= 2.0)	0.0	-0.8
2	Power Exponential Distribution (α= 0.6)	0.0	2.2
3	Power Method Distribution (b=1.112, c=0.174, d=-0.050)	0.75	0.0
4	Power Method Distribution (b=0.936, c=0.268, d=-0.004)	1.50	3.0
5	Normal Distribution	0.0	0.0
6	Rayleigh Distribution (β= 1)	0.631	0.245

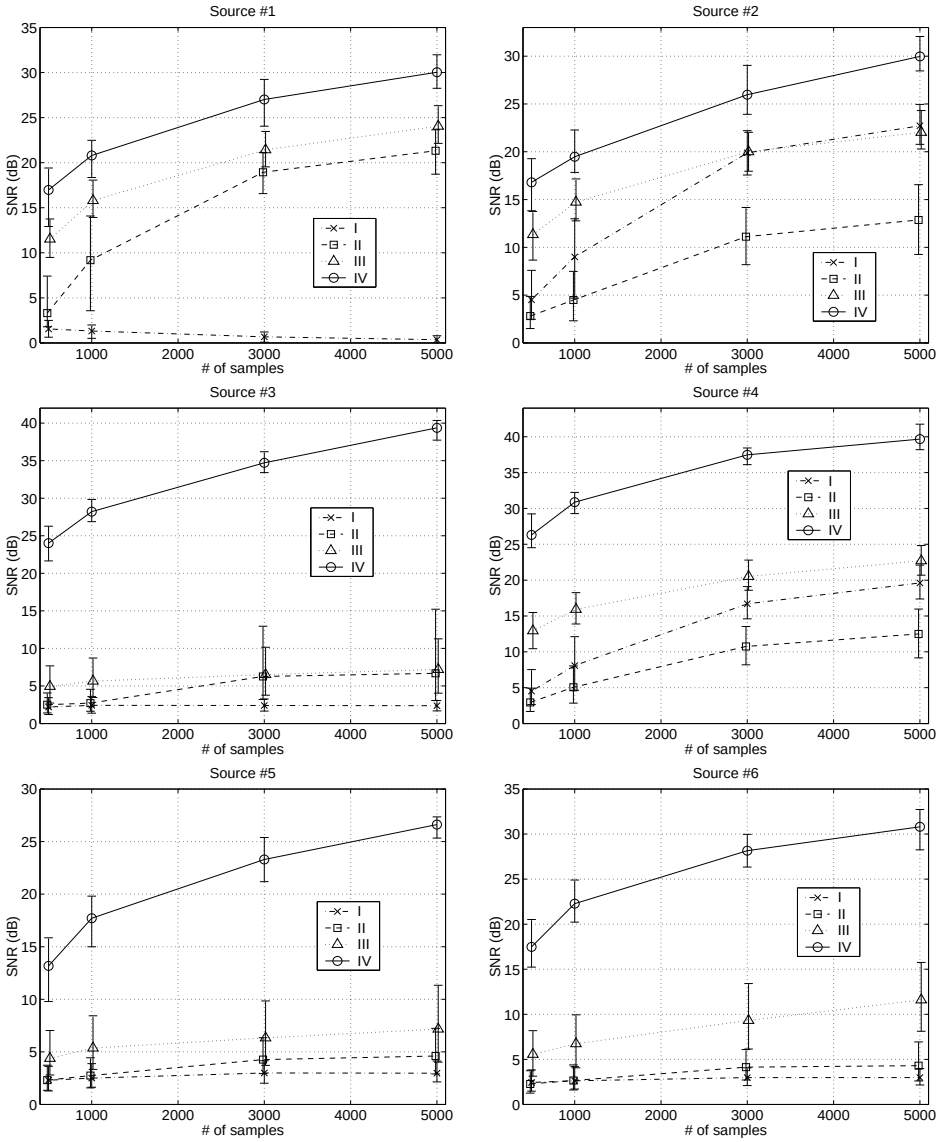

Fig. 1. First simulation experiment: *I-InfoMax ICA, II-Extended InfoMax ICA, III-Jade, IV-Non-parametric ICA*. The separation results for the six different sources of Table 1 are averaged over 1000 simulations. The accuracy of the separation is measured in terms of median log signal-to-noise ratio (SNR), defined as $10 \log_{10} \sum_{m=1}^{M} s_m^2 / \sum_{m=1}^{M} (s_m - \hat{s}_m)^2$ (dB), where $\mathbf{s}$ is the original signal and $\hat{\mathbf{s}}$ is the reconstructed signal. For standard sub-gaussian or super-gaussian sources (#1, #2, and #4) the non-parametric ICA outperforms the other methods. Moreover, it is capable of accurately separating sources #3, #5, and #6, which the other algorithms fail to separate. The vertical bars extend between the 25% and the 75% percentiles.

Table 2. Second simulation experiment. The performance in separating mixtures of a gaussian source with one of five types of skewed sources (kurtosis ≈ 0.75) is investigated. The median SNR (in dB) along with 25% and 75% percentiles are reported for 100 runs (averaged over the two sources). For increasing skewness values, the non-parametric ICA clearly shows a substantial performance improvement, while the other ICA algorithms present a degradation in the separation performance.

	skewness = 0			skewness = 0.25			skewness = 0.5			skewness = 0.75			skewness = 1.0		
	$\hat{\text{snr}}$	25%	75%	$\hat{\text{snr}}$	25%	75%	$\hat{\text{snr}}$	25%	75%	$\hat{\text{snr}}$	25%	75%	$\hat{\text{snr}}$	25%	75%
Original InfoMax	17.0	10.5	25.1	15.9	9.7	21.9	13.6	9.3	21.7	13.4	7.7	18.8	10.5	6.5	17.2
Extended InfoMax	10.0	5.6	16.3	8.5	5.3	15.1	8.0	5.3	14.1	8.4	5.3	14.9	7.4	5.1	13.8
Jade	18.7	15.2	23.0	17.6	13.7	22.8	16.9	13.1	21.8	15.3	11.3	19.2	15.4	12.3	20.0
Non-Parametric ICA	**19.2**	13.8	25.0	**24.4**	18.9	28.8	**25.7**	18.4	31.8	**33.0**	27.4	36.7	**42.7**	36.8	47.6

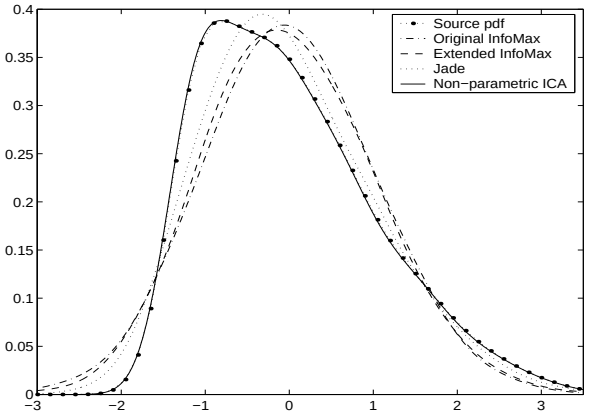

Fig. 2. Pdf of source#3 estimated from the signals reconstructed by each BSS algorithm. Only the proposed algorithm accurately estimates the pdf of this source.

of signals accurately. Our results show that the proposed method is able to achieve the required quality of separation (20dB) with only 750 samples, while the Extended InfoMax ICA and Jade require almost twice as much data samples.

5. CONCLUSIONS

We introduced a novel non-parametric blind signal separation technique based on a true Projection Pursuit framework. Simulation results show that the proposed approach is capable of separating mixtures of a broad class of signals, with a noticeable performance improvement when compared to existing ICA algorithms. The ability of separating arbitrarily distributed sources, combined with favorable convergence properties, and a relatively modest computational complexity, establish the non-parametric ICA algorithm as an attractive alternative to existing ICA methods. We would like to thank Dr. Lieven Vandenberghe for his valuable suggestions on the optimization methods.

6. REFERENCES

[1] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[2] H.B. Barlow, "Unsupervised learning," *Neural Computation*, vol. 1, pp. 295–311, 1989.

[3] Aapo Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999.

[4] Te-Won Lee, Mark Girolami, and Terrence J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 417–441, 1999.

[5] Jean-François Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan. 1999.

[6] J. Karvanen J. Eriksson and V. Koivunen, "Source distribution adaptive maximum likelihood estimation of ica model," in *Proceedings of ICA 2000*, P. Pajunen and J. Karhunen Editors, Eds., Helsinki, 2000, pp. 227–232.

[7] N. Vlassis and Y. Motomura, "Efficient source adaptivity in independent component analysis," *IEEE Trans. Neural Networks*, vol. 12, no. 3, pp. 559–566, May 2001.

[8] M.C. Jones, *The Projection Pursuit Algorithm for Exploratory Data Analysis*, Ph.D. thesis, University of Bath, School of Mathematics, 1983.

[9] J. Friedman and J.W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. on Computers*, vol. C-23, no. 9, pp. 881–889, 1974.

[10] P.J. Huber, "Projection pursuit," *Annals of Statistics*, vol. 13, Issue 2, pp. 435–475, June 1985.

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

[12] M. Girolami and C. Fyfe, "An extended exploratory projection pursuit network with linear and non-linear anti-hebbian lateral connections applied to the cocktail party problem," *Neural Networks*, vol. 10, no. 9, pp. 1607–1618, 1997.

[13] D. Obradovic and G. Deco, "Information maximization and independent component analysis: Is there a difference?," *Neural Computation*, vol. 10, pp. 2085–2101, 1998.

[14] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 1985.

[15] Allen I. Fleishman, "A method for simulating non-normal distributions," *Psychometrika*, vol. 43, no. 4, pp. 521–532, Dec. 1978.