PEARSON SYSTEM BASED METHOD FOR BLIND SEPARATION

Juha Karvanen, Jan Eriksson*, and Visa Koivunen*

Signal Processing Laboratory
Helsinki University of Technology
P.O. Box 3000, FIN-02015 HUT, Finland
{juha.karvanen, jan.eriksson, visa.koivunen}@hut.fi

ABSTRACT

In this paper a new approach for Blind Source Separation (BSS) is introduced. The Pearson system is employed for modeling source distributions. The major benefit of the Pearson system is that it also takes into account the skewness of the distributions. We briefly review the Pearson system and study how its parameters may be estimated. A BSS method combining Pearson system based on the maximum likelihood approach and fixed contrast functions is presented and algorithms for its maximization are proposed. The simulation examples illustrate that the proposed method reliably separates the sources in situations where some widely used BSS methods may perform poorly.

1. INTRODUCTION

In this paper we introduce a new method for Blind Source Separation (BSS). The method is applicable to a wide class of source distributions that may also be skewed and may even have zero kurtosis. Such sources occur, for instance, in biomedical signals and communications. The underlying source distributions are modeled using the Pearson system (see [10]).

We consider the classical ICA model with instantaneous mixing

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

where the sources $\mathbf{s} = [s_1, s_2, \dots, s_m]^T$ are mutually independent random variables and $\mathbf{A}_{m \times m}$ is an unknown invertible mixing matrix. The goal is to find only from the observations, $\mathbf{x}$, a matrix $\mathbf{W}$ such that the output

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{2}$$

is an estimate of the possibly scaled and permutated source vector $\mathbf{s}$.

* This work was funded by the Academy of Finland.

The proposed method combines two well-known techniques for Independent Component Analysis (ICA): fixed non-linear contrast functions and maximum likelihood approach. In the maximum likelihood approach we use Pearson system to model source distributions. The Pearson system covers an extensive range of different values of kurtosis and skewness and includes many distributions with practical importance. The rational of Pearson system in BSS is that the model is flexible enough to adapt to different source distributions that may also be asymmetric. The fixed contrast functions improve the speed and stability of the Pearson system based method in the cases where sources are easily separable.

This paper is organized as follows: in Section 2 the Pearson system is reviewed. Estimation of its parameters is considered as well. Section 3 introduces a new BSS method where the Pearson system is employed. In particular new objective function is derived and algorithms for optimizing it are considered. Finally, simulation examples are given in Section 4.

2. THE PEARSON SYSTEM

The Pearson system is a parametric family of distributions that may be used to model a wide class of source distributions. The Pearson system has a great importance in statistics and it has been extensively studied; see for instance [9] for references. The Pearson system is defined by the differential equation

$$f'(x) = \frac{(x-a)f(x)}{b_0 + b_1 x + b_2 x^2}, \tag{3}$$

where a, b_0, b_1 and b_2 are the parameters of the distribution. In the maximum likelihood approach to ICA the score function of hypothesized source distribution is used as a contrast. The score function of the Pearson system is easily solved from equation (3)

$$\varphi(x) = -\frac{f^{'}(x)}{f(x)} = -\frac{(x-a)}{b_0 + b_1 x + b_2 x^2}. \qquad (4)$$

The simplicity of score makes Pearson system particularly appealing for ICA. For the derivative of the score function we have

$$\varphi^{'}(x) = -\frac{b_0 + a\, b_1 + 2\, a\, x\, b_2 - x^2\, b_2}{(b_0 + b_1 x + b_2 x^2)^2}. \qquad (5)$$

The parameters a, b_0, b_1 and b_2 may be estimated by the method of moments. The correspondence between parameters a, b_0, b_1 and b_2 and second μ_2, third μ_3 and fourth μ_4 central moment of distribution is the following [10]

$$b_1 = a \quad = \quad -\frac{\mu_3(\mu_4 + 3\mu_2^2)}{C} \qquad (6)$$

$$b_0 \quad = \quad -\frac{\mu_2(4\mu_2\mu_4 - 3\mu_3^2)}{C} \qquad (7)$$

$$b_2 \quad = \quad -\frac{(2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3)}{C}, \qquad (8)$$

where $C = 10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3$.

In the method of moments, theoretical moments are estimated by sample moments

$$\hat{\alpha}_1 \quad = \quad \bar{x} = \sum_{i=1}^{n} x_i/n \qquad (9)$$

$$\hat{\alpha}_2 \quad = \quad \hat{\sigma}^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/n \qquad (10)$$

$$\hat{\alpha}_3 \quad = \quad \sum_{i=1}^{n} (x_i - \bar{x})^3/(n\hat{\sigma}^3) \qquad (11)$$

$$\hat{\alpha}_4 \quad = \quad \sum_{i=1}^{n} (x_i - \bar{x})^4/(n\hat{\sigma}^4) \qquad (12)$$

computed from data. When mean is zero and variance is one the following formulas are obtained for estimators of Pearson system parameters

$$\hat{b_1} = \hat{a} \quad = \quad -\frac{\hat{\alpha}_3(\hat{\alpha}_4 + 3)}{\hat{C}} \qquad (13)$$

$$\hat{b_0} \quad = \quad -\frac{(4\hat{\alpha}_4 - 3\hat{\alpha}_3^2)}{\hat{C}} \qquad (14)$$

$$\hat{b_2} \quad = \quad -\frac{(2\hat{\alpha}_4 - 3\hat{\alpha}_3^2 - 6)}{\hat{C}}, \qquad (15)$$

where $\hat{C} = 10\hat{\alpha}_4 - 12\hat{\alpha}_3^2 - 18$. It is seen that $b_1 = a$ so the number of parameters actually reduces to three.

When the denominator in (4) have two real roots the Pearson system represent generalized beta distribution. Since the beta distribution is defined in a finite interval the method of moments estimator may lead to the model where some observations are outside of the definition interval. A viable solution to this problem is to exploit sample minimum and maximum in estimation.

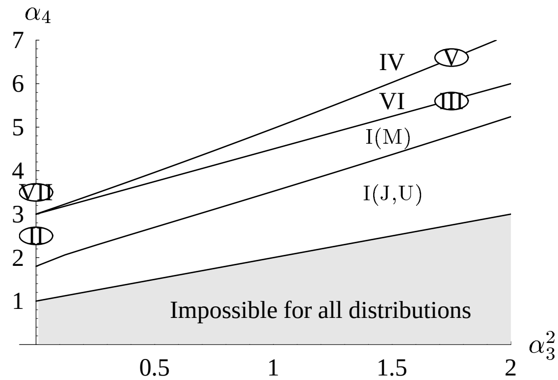


Figure 1: An illustration of the Pearson system in (α_3^2, α_4)-plane. Limit for all distributions is line $\alpha_4 = \alpha_3^2 + 1$. The Latin numbers refer to the traditional classification of Pearson distributions. Types I and II are beta distributions of first kind. Notation I(J,U) refers to J- and U-shaped distributions and I(M) to unimodal distribution. The boundary between I(J,U) and I(M) is curve $4(4\alpha_4 - 3\alpha_3^2)(5\alpha_4 - 6\alpha_3^2 - 9)^2 = \alpha_3^2(\alpha_4 + 3)^2(8\alpha_4 - 9\alpha_3^2 - 12)$ Type III is Gamma distribution for which $\alpha_4 = \frac{3}{2}\alpha_3^2 - 3$. Type VI is the beta distribution of second kind. Type V is characterized by curve $\alpha_3^2(\alpha_4 + 3)^2 = 4(4\alpha_4 - 3\alpha_3^2)(2\alpha_4 - 3\alpha_3^2 - 6)$. Type IV is the case where the equation $b_0 + b_1 + b_2 x^2 = 0$ has complex roots. Type VII is the Student's t-distribution.

Many widely used distributions, including normal, Student's t, gamma and beta distribution belong to the Pearson family. This is illustrated in Figure 1.

3. PEARSON SYSTEM BASED ICA

It is shown [3] that if the source distributions are known, the score functions are the optimal choice for the contrast function. In the maximum likelihood approach source distributions are estimated by a parametric model. In the method we propose, the underlying source distributions are estimated through the marginal distributions by fitting them to the Pearson family using the method of moments as described in the previous section. Fitting to the Pearson system is done iteratively until the optimization algorithm converges.

The actual algorithm optimizing the derived criterion could be any suitable ICA algorithm where maximum likelihood contrasts are utilized, such as natural gradient [1] or relative gradient [3] algorithm

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \eta \left(I - \varphi(\mathbf{y})\mathbf{y}^T \right) \mathbf{W}_k, \qquad (16)$$

where η is the learning rate, or fixed point algorithm [6, 7]

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mathbf{D} \left(E\{\varphi(\mathbf{y})\mathbf{y}^T\} - \mathrm{diag}(E\{\varphi(y_i)y_i\}) \right) \mathbf{W}_k, \qquad (17)$$

where $D = \mathrm{diag}\left(1/(E\{\varphi(y_i)y_i\} - E\{\varphi^{'}(y_i)\}\right)$.

Since the Pearson system excels in modeling distributions that are close to normal distribution but does not offer any remarkable advantages in modeling distributions that are far from normal distribution, we employ widely used fixed non-linearities as a contrast function when the kurtosis greatly differ from the kurtosis of the normal distribution. By this approach we can speed up the computation and avoid the estimation problems that might arise when a distribution is strongly heavy tailed or J-shaped or U-shaped distribution on a finite interval. The fixed non-linearities can be any suitably chosen non-linearities. For instance, we may use cubic contrast when distribution is clearly sub-Gaussian, the Pearson system when the distribution is nearly Gaussian and hyperbolic tangent contrast when distribution is clearly super-Gaussian. The benefit of using Pearson system is that the method can separate sources with a skewed distribution and the same kurtosis as the Gaussian distribution.

In our experiments we used fixed point algorithm and hyperbolic tangent contrast $\varphi(y) = \tanh(2y)$ for both clearly sub-gaussian and clearly super-gaussian sources. The boundaries between contrasts are presented in Figure 2. The procedure for the Pearson-ICA may be given as follows:

Repeat until convergence [1]

1. Calculate the third and fourth sample moments α_3 and α_4 for current data $\mathbf{y}_k = \mathbf{W}_k\mathbf{x}$ and select the Pearson system or fixed (tanh) contrast according to Figure 2.

2. If the Pearson system was selected estimate parameters of the distribution by method of moments.

3. Calculate scores $\varphi(\mathbf{y}_k)$ for the Pearson system or fixed contrast.

4. Calculate the demixing matrix $\mathbf{W}_{k+1}$ using algorithm (16) or (17).

[1]The convergence criterion can be any suitable for the gradient algorithms and the fixed-point algorithm respectively. In our experiments, we used a criterion similar to the symmetric FastICA[5] with $\varepsilon = 0.0001$ in order to make results comparable.

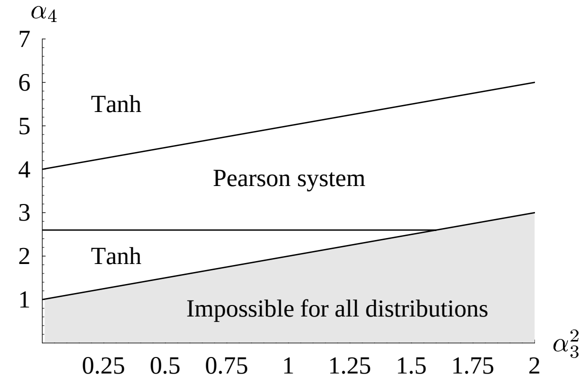


Figure 2: The contrast functions used in the Pearson-ICA are presented in (α_3^2, α_4)-plane. Limit for all distributions is line $\alpha_4 = \alpha_3^2 + 1$. Clearly sub-gaussian sources are defined to have kurtosis less than 2.6 and the tanh-contrast is utilized. Clearly super-gaussian sources are defined to have $\alpha_4 = \alpha_3^2 + 4$ and the tanh-contrast is again utilized. In area between these boundaries the Pearson system is used. The choice of boundaries is based on practical experience.

Since both moment estimators for parameters and score function are simple rational functions, the Pearson-ICA is computationally fast. If sources are nearly Gaussian it appears to be faster than fixed non-linearities, probably because maximum likelihood contrast provides faster convergence. We may even combine parameter estimation (13) and score calculation (4) to achieve the score directly as function of sample moments. After some manipulations we obtain

$$\varphi(x) = \left[\alpha_3\left(3+\alpha_4\right) - x\left(12{\alpha_3}^2 - 5\alpha_4 + 18\right)\right]/B \qquad (18)$$

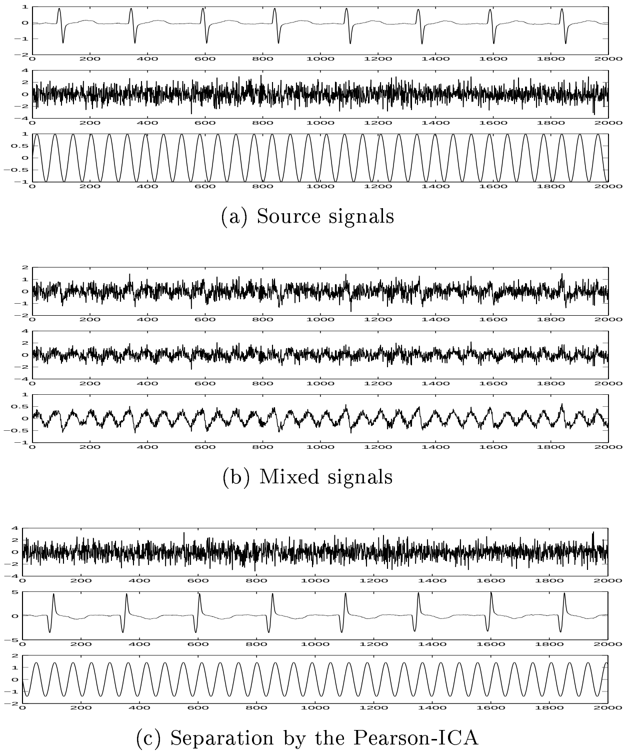
and

$$\begin{aligned}\varphi^{'}(x) &= \frac{-45{\alpha_3}^2 - 36{\alpha_3}^4 + 84{\alpha_3}^2\alpha_4 + {\alpha_3}^2{\alpha_4}^2 + 72\alpha_4 - 40{\alpha_4}^2}{B^2} \\ &+ \frac{x^2\left(126{\alpha_3}^2 + 36{\alpha_3}^4 - 54{\alpha_3}^2\alpha_4 + 20\alpha_4^2 - 96\alpha_4 + 108\right)}{B^2} \\ &+ \frac{x\left(-6{\alpha_3}^3\left(3+\alpha_4\right) + 4\alpha_3\left(-9+{\alpha_4}^2\right)\right)}{B^2},\end{aligned} \qquad (19)$$

where $B = 3{\alpha_3}^2 - 4\alpha_4 - x\alpha_3\left(3+\alpha_4\right) + x^2\left(6 + 3{\alpha_3}^2 - 2\alpha_4\right)$.

4. SIMULATION EXAMPLES

In order to illustrate the performance of the proposed algorithm, we consider first an example with a mixture of three sources: a sine wave (sub-Gaussian), a synthetic ECG signal (super-Gaussian), and a random Gaussian sequence with zero mean and unit variance. This is an easy case: all common ICA methods



(a) Source signals

(b) Mixed signals

(c) Separation by the Pearson-ICA

Figure 3: Separation of a sine wave, an ECG signal, and a Gaussian noise signal with Pearson-ICA.

		Sample size		
Source	Method	2000	5000	10000
ECG	Pearson	24.98	34.27	36.46
	Pow3	22.75	29.32	31.17
	Tanh	25.89	35.08	37.20
	Gauss	26.63	36.52	38.81
	JADE	23.54	29.01	30.77
Sine	Pearson	13.03	40.51	44.92
	Pow3	15.03	31.14	35.19
	Tanh	14.18	39.56	43.60
	Gauss	14.02	42.16	41.96
	JADE	14.91	32.96	37.38
Gaussian	Pearson	28.19	33.11	36.15
	Pow3	28.13	32.76	35.62
	Tanh	30.37	34.57	37.75
	Gauss	30.55	34.76	37.95
	JADE	30.99	35.30	38.49

Table 1: The Signal to Interference Ratio (SIR(dB)$=-10\log_{10}$(MSE)) error performance of Pearson-ICA, JADE and different contrast functions of FastICA (symmetric approach). The sources and separated signals were normalized to zero mean and unit variance. The SIR values are averages over 1000 realizations.

are expected to perform very well. The Pearson-ICA is compared with three different non-linearity in FastICA package [5] and JADE-algorithm [2]. Three signal lengths, 2000, 5000 and 10000 are tried generating 1000 realizations from each. Non-singular mixing matrix was randomly generated for each of 3000 simulation. The Figure 3 illustrates source signals, mixture and separation result by Pearson-ICA. It can be seen that the sources are well separated. The comparison of methods is done computing the Signal to Interference Ratios (SIR(dB)$= -10\log_{10}$(MSE)) between the zero mean, the unit variance normalized signals. The sources and the sign adjusted signals are matched by taking the signal with the minimum SIR value to be the separation estimate. The average SIR values for different signal lengths are shown in Table 1. The conclusion is that there are no significant differences in separation performance between Pearson-ICA and the four other algorithms used in the comparison.

In the second example we demonstrate that the Pearson-ICA algorithm can separate signals with kurtosis equal to that of Gaussian distribution. Again 1000 realizations of seven different signal lengths 100, 200, 500, 1000, 2000, 5000 and 10000 are generated. The theoretical distributions of the signals are: Lognormal(0.1,0.15), Rayleigh(1), Normal(0,1), and GLD(0.2370,0.1983,0.1672,0.1065). The acronym GLD refers to Generalized Lambda Distribution [8, 4] that is an extensive four parametric family of distributions defined by the inverse distribution function. The GLD(0.2370,0.1983,0.1672,0.1065) distribution has the theoretical moments $\alpha_1 = 0$, $\alpha_2 = 1$, $\alpha_3 = 0.2$, and $\alpha_4 = 3$. Lognormal and Rayleigh distributions are commonly used for modeling the fading communication channels. The sources are mixed using the (randomly generated) matrix

$$\mathbf{A} = \begin{pmatrix} 0.7396 & 0.9084 & 0.2994 & 0.3089 \\ 0.4898 & 0.2980 & 0.5771 & 0.4108 \\ 0.1096 & 0.7808 & 0.8361 & 0.4669 \\ 0.4199 & 0.8799 & 0.2706 & 0.7467 \end{pmatrix}. \quad (20)$$

The first hundred observations from one realization of the Pearson-ICA and kurtosis separated signals are plotted in Figure 4. The SIR averages are presented in Figure 5. As it can be seen the Pearson-ICA performs well for all the sources. The FastICA algorithm with the fixed contrast functions is unable to separate the normal signal from the GLD signal. This also reflects to the SIR values of the other two signals. JADE algorithm perform slightly better than FastICA algorithm but is outperformed by Pearson-ICA. The results indicate that in this case about 1000 observations are needed in order to achieve good separation by Pearson-ICA. Naturally, this result depends on source distributions, and an estimators used for Pearson parameters.

5. CONCLUSION

We introduced a Pearson System based algorithm for estimating ICA model. The proposed method combines the best features of fixed non-linearities and maximum likelihood approach. Source distribution adaptive contrast function is derived using the Pearson system as the model. We have presented the theoretical background for the use of Pearson-ICA, and shown that it can separate a wide class of source signals including sub- and super-Gaussian, and even skewed distributions with zero kurtosis.

6. REFERENCES

[1] S.-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10:251–276, 1998.

[2] J. F. Cardoso. Code with references available at `http://sig.enst.fr:80/~cardoso/`.

[3] J. F. Cardoso. Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, 1998.

[4] J. Eriksson, J. Karvanen, and V. Koivunen. Source distribution adaptive maximum likelihood estimation of ICA model. Submitted for publication in ICA2000.

[5] A. Hyvärinen. Code with references available at `http://www.cis.hut.fi/projects/ica/fastica/`.

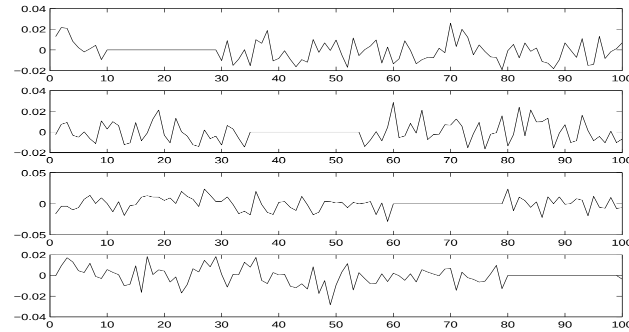
[6] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[7] A. Hyvärinen. The fixed-point algorithm and maximum likelihood estimation for independent component analysis. *Neural Processing Letters*, 10(1):1–5, 1999.

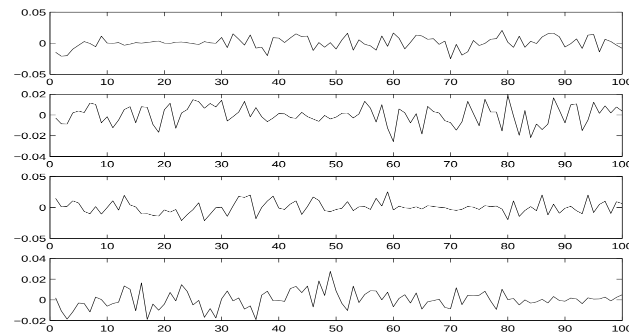
[8] Z. A. Karian, E. J. Dudewicz, and P. McDonald. The extended generalized lambda distribution system for fitting distributions to data: history, completion of theory, tables, applications, the "final word" on moment fits. *Comm. Stat.: Simulation and Computation*, 25(3):611–642, 1996.

[9] J. K. Ord. *Families of Frequency Distributions.* Griffin, London, 1972.

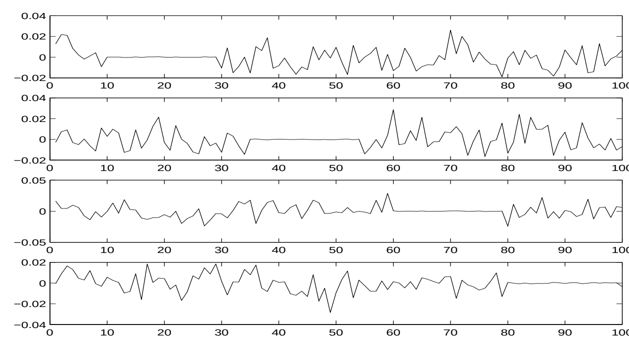
[10] A. Stuart and J. K. Ord. *Kendall's advanced theory of statistics: Distribution theory*, volume 1. Edward Arnold, sixth edition, 1994.



(a) Source signals

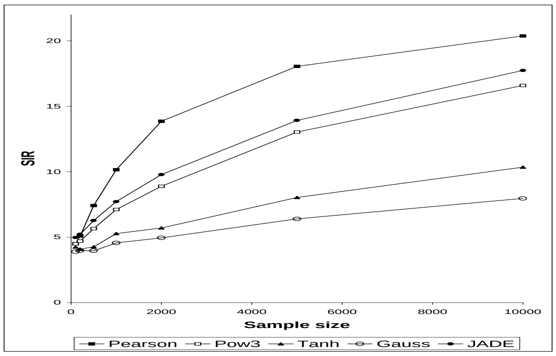


(b) Separation by kurtosis (FastICA:Pow3)

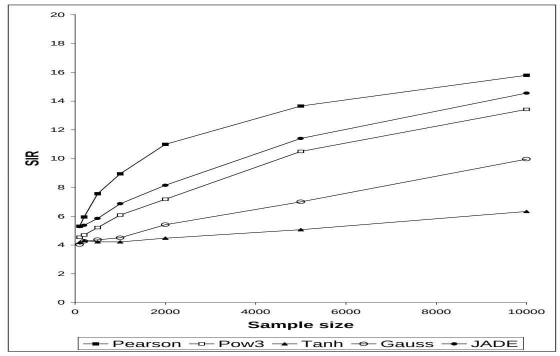


(c) Separation by the Pearson-ICA

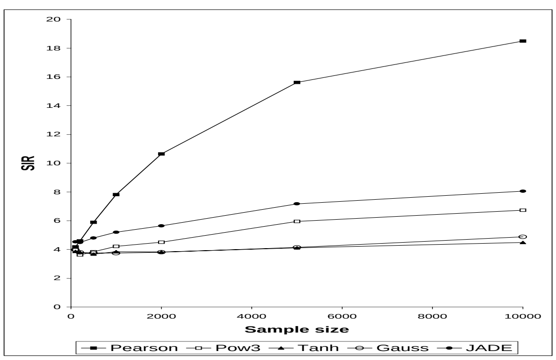
Figure 4: Separation results of the Pearson-ICA and the kurtosis criteria are compared in the case where the sources are Lognormal(0.1,0.15), Rayleigh(1), Normal(0,1), and GLD(0.2370,0.1983,0.1672,0.1065). The number of samples was 10000 but only the first 100 observations are shown. In order to visualize the quality of separation a short sequence of zeros was added to every source signal. It can be seen that the kurtosis based method fails to separate the GLD and the normal signals, but the Pearson-ICA reliably separates all the sources.



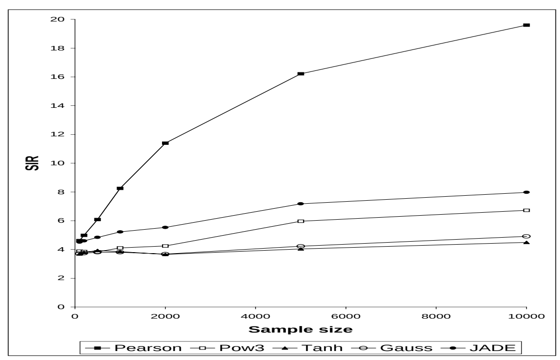
(a) Lognormal distributed source



(b) Rayleigh distributed source



(c) Normal distributed source



(d) GLD distributed source

Figure 5: The Signal to Interference Ratio (SIR(dB)=$-10\log_{10}$(MSE)) error performance of Pearson-ICA and FastICA contrasts Pow3, Tanh and Gauss and JADE algorithm as the function of the sample size for sources generated from Lognormal(0.1,0.15), Rayleigh(1), Normal(0,1) and GLD(0.2370,0.1983,0.1672,0.1065). The sources and separated signals were normalized to zero mean and unit variance. The SIR values are averages over 1000 realizations.