BLIND SOURCE SEPARATION USING MAXIMUM ENTROPY PDF ESTIMATION BASED ON FRACTIONAL MOMENTS

Babak Abbasi Bastami and Hamidreza Amindavar

Amirkabir University of Technology, Department of Electrical Engineering, Tehran, Iran

Abstract. Recovering a set of independent sources which are linearly mixed is the main task of the blind source separation. Utilizing different methods such as infomax principle, mutual information and maximum likelihood leads to simple iterative procedures such as natural gradient algorithms. These algorithms depend on a nonlinear function (known as score or activation function) of source distributions. Since there is no prior knowledge of source distributions, the optimality of the algorithms is based on the choice of a suitable parametric density model. In this paper, we propose an adaptive optimal score function based on the fractional moments of the sources. In order to obtain a parametric model for the source distributions, we use a few sampled fractional moments to construct the maximum entropy probability density function (PDF) estimation . By applying an optimization method we can obtain the optimal fractional moments that best fit the source distributions. Using the fractional moments (FM) instead of the integer moments causes the maximum entropy estimated PDF to converge to the true PDF much faster . The simulation results show that unlike the most previous proposed models for the nonlinear score function, which are limited to some sorts of source families such as sub-gaussian and super-gaussian or some forms of source distribution models such as generalized gaussian distribution, our new model achieves better results for every source signal without any prior assumption for its randomness behavior.

INTRODUCTION

The goal of Blind Source Separation (BSS) is to recover unknown independent source signals from the observations. We consider a linear mixture of source vector as $\mathbf{x} = A\mathbf{s}$, where $\mathbf{s}$, is the vector of the source signals, A, is the mixing matrix and $\mathbf{x}$, is the vector of the observed signals. We assume that the number of sources and the sensors are equal. The aim is to recover the independent source signals, using a linear de-mixing model $\mathbf{y} = W\mathbf{x}$. The statistical approaches for solving the problem such as maximum likelihood, InfoMax algorithm, Kullback-Leibler divergence and ..., all would eventually lead to some gradient based learning rules for updating iteratively the de-mixing matrix W. The main problem in these learning rules, is to find an optimal stable choice of the nonlinear score or activation function for various distributions. They commonly use a fixed but nonlinear activation function resulting in an acceptable performance, however, just for some specific sorts of source densities. In most BSS problems, the source distributions are classified as near-gaussian, sub-gaussian and super-gaussian models . The well-known nonlinearities can only be applied for one of these density models. It can be proven that [1] there exist adaptive nonlinear activation functions satisfying the stability condition of the learning rule and suitable for every source distribution model. In this paper, we propose an adaptive score function by means of approximating the den-

sity function applying Shannon's maximum entropy principle using fractional moments. Unlike the previous common used nonlinearities, the proposed adaptive activation function can perform efficiently in every signal situation regardless of its distribution. If a set of moments meet the Carleman condition [2] then a unique PDF can be determined out of them. Moments are attractive because their computation is algorithmically simple and uniquely defined for any signal; it can be carried out in parallel and therefore very fast, and, since moments are global quantities, all available information is used making moment-based methods less vulnerable to losses or changes of details than methods that use few particular features of the signal. However, moments become very noise-sensitive with increasing order. Hence, the lowest possible orders should be used in a moment-based procedure. The classical moment based methods involves very few integer moments. We describe how a very few fractional- and possibly negative-order moments can be used to increase the accuracy of PDF estimation in MAXENT sense in a generally distributed sources . However, all fractional moments may not equally be suitable for estimating PDF of the signal. In this paper, we estimate the PDF of a signal adaptively using MAXENT method using the optimum FM. In our scheme we use MAXENT method which has been involved in the solution of many statistical problems and we use it to fit the PDF for signal. The chief assertion of the MAXENT PDF estimation is that the most unbiased PDF is the maximum entropy distribution satisfying some constraints which are usually a number of known moments, fractional or integer. The best way to give a short introduction to MAXENT is to offer a quote from one of the pioneers of these techniques, Edwin Jaynes [3]: The notion of entropy defines a kind of measure on the space of probability distributions, such that those of high entropy are in some sense favored over others. The maximum entropy distributions are "in some sense favored" can be backed up by mathematically proving what has come to be called the concentration theorem [3]. The result of this is that for a given set of constraints such as moments or their functions, if there is a family densities that could give us our solution, most of the solutions are concentrated, or close to the maximum entropy PDF. Thus, it is our best guess to take MAXENT PDF as the distribution of the desired variate. This paper is organized as follows. In section 2, a discussion about MAXENT PDF estimation based on FM is provided, in section 3 we discuss the adaptive learning rules in BSS and the well-known activation functions, in section 4 the new adaptive nonlinear score function based on MAXENT fractional moment PDF estimation will be proposed, in section 5 we provide the simulations of an experiment and results and some concluding remarks at the end.

DENSITY ESTIMATION VIA FRACTIONAL MOMENTS

It is a well known fact that a finite set of moments does not allow to calculate PDF of a random process. To get an unambiguous statistic, one has to approximate the unspecified moments in some sense. One way to do this is maximization of differential entropy. In this paper, we utilize the MAXENT principle as follows. Given the received samples of a signal, we estimate PDF in MAXENT sense that matches the received data. We note that the traditional MAXENT [3] approach is based on a give set of moments or estimated sample moments to estimate PDF in MAXENT sense, but in this

paper, we find the best set of moments, fractional or integer, that fit the received data set optimally in MAXENT sense. It is shown that MAXENT PDF estimation based on fractional moments has better performance than integer moments [4, 5]. We consider a positive random variable X with PDF $f(x)$. Our problem is to maximize the entropy functional $H[f] = -\int_0^\infty f(x) \ln f(x) dx$ subject to some FM $\mu_j = \{E(X^{\alpha_j})\}_{j=0}^M$ where the FM based MAXENT PDF is given as follows

$$f_M(x) = \exp\left(-\sum_{j=0}^M \lambda_j x^{\alpha_j}\right), \tag{1}$$

where $\lambda_0, \cdots, \lambda_M$ are the Lagrange multipliers corresponding to the following M FM constraints

$$\mu_{\alpha_j} = E(X^{\alpha_j}) = \int_0^\infty x^{\alpha_j} f_M(x) dx, \quad j = 0, \cdots, M, \tag{2}$$

where $\alpha_0 = 0$. Then the entropy is represented by

$$H[f_M] = -\int_0^\infty f_M(x) \ln f_M(x) dx = \sum_{j=0}^M \lambda_j \mu_{\alpha_j}. \tag{3}$$

If we assume that $F(x)$ and $F_M(x)$ are the cumulative distribution function for the exact and MAXENT solution, respectively, it has been shown [4, 5, 6] that we have the following bound for the difference between these two functions

$$\sup_{x \in [0,\infty)} |F_M(x) - F(x)| \le 3\sqrt{-1 + \sqrt{1 + \frac{4}{9}(H[f_M] - H[f])}},$$

therefore, a convergence in entropy is translated into convergence in distribution. If we define the divergence measure of two PDF's as $\int_0^\infty f(x) \ln (f(x)/f_M(x)) dx$, whenever the two PDF's have the same fractional moments we have

$$\int_0^\infty f(x) \ln \frac{f(x)}{f_M(x)} dx = H[f_M] - H[f] \tag{4}$$

Hence the two entropies converge to each other in the case of the FM's equivalence. Therefore we can always find an optimal choice for the fractional parameters α_j [4, 5, 6]. We assume $\{x_1, \cdots, x_N\}$ are the received samples, then, in order to determine the parameters of $f_M(x)$ in (1), we implement the following optimization for $j = 0, \cdots, M$

$$\min_{\alpha_j, \lambda_j} H[f_M] = \sum_{j=0}^M \lambda_j \hat{\mu}_{\alpha_j}, \quad \hat{\mu}_{\alpha_j} = \frac{1}{N} \sum_{n=1}^N x_n^{\alpha_j}. \tag{5}$$

Also, it has been proven that the convergence to the exact PDF holds as $M \to \infty$ [5, 6]. Our optimization results show that applying FM instead of the integer moments causes the MAXENT estimated PDF $f_M(x)$ to converge to $f(x)$ much faster. For example in Figure 1, we compare the resulting MAXENT density estimates by 5000 samples of a

Gamma distributed random variable with PDF $f(x) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b}$. Using the first four sample integer moments, and two sample FM, that the optimization (5) determines (0.8392,1.9289), we arrive at the following estimates for the PDF of Gamma distribution with parameter $a = 1, b = 1$ for $x \in [x_{\min} = 0.0001, x_{\max} = 10.2]$

$$f_M(x) = \exp\left(-0.2155 - 0.5030x - 0.3150x^2 + 0.0718x^3 - 0.0053x^4\right),$$
$$f_M(x) = \exp\left(0.0984 - 1.10965x^{0.8392} - 0.0362x^{1.9289}\right)$$

Our comparative measure is the relative error defined as

$$\text{relative error} = \frac{|\text{True PDF} - \text{Approximated PDF}|}{\text{True PDF}}.$$

As it is shown in Figure 1, the PDF obtained via two optimized FM provide a better accuracy over the MAXENT PDF estimator based on four integer moments.

ADAPTIVE LEARNING ALGORITHMS IN BLIND SOURCE SEPARATION

Several information theoretic approaches yield the adaptive learning algorithms such as[7]
$$\mathbf{W_{j+1}} = \mathbf{W_j} + \mu(\mathbf{I} - \mathbf{g(y)y}^T)\mathbf{W_j} \tag{6}$$
for blind source separation and independent component analysis (ICA) problems. In equation 6, μ is the learning rate and $\mathbf{g(y)}$ is known as the score or activation function, it is a vector of nonlinearities dependent on the distribution functions of the observed output signals. Instead of equation 6, the adaptive fixed point algorithms without any learning rule can also be utilized.

The stability condition in the learning rule 6, depends on the form of the nonlinearity $\mathbf{g}$. For local stability, the signal must satisfy the equation [8]

$$E\{\mathbf{g}'(\mathbf{y})\}E\{\mathbf{y}^2\} - E\{\mathbf{g(y)y}\} > 0. \tag{7}$$

After some manipulations, the optimal nonlinear score function will be obtained as [1]

$$\mathbf{g(y)} = -\frac{\mathbf{f}'(\mathbf{y})}{\mathbf{f(y)}} \tag{8}$$

where $\mathbf{f}$ is the probability distribution function of the observed signals. The above equation can also be obtained by other classical methods such as Maximum Likelihood or InfoMax [7]. The optimal activation function requires a prior knowledge of source distributions which are usually unavailable. It has been proven in [1] that it is impossible to find a fix nonlinearity suitable for all sorts of distributions, but there exists an adaptive nonlinearity which can be suitable for all kinds of densities by changing the nonlinearity function variables. There are some density models that lead to some nonlinear score functions, which are widely used in ICA algorithms. These widely used nonlinearities

are usually appropriate choices for just some specific kinds of sources but they can not be considered as suitable options for the other sorts of source distributions. For instance, the fix cubic nonlinear activation function, $\mathbf{g}(\mathbf{y}) = \mathbf{y}^3$, has been known as an appropriate choice for sub-Gaussian signals. The hyperbolic-Cauchy score function model, $\mathbf{g}(\mathbf{y}) = \tanh(\gamma\mathbf{y})$, where $\gamma = 1/\sigma_y^2$ is another commonly used nonlinearity which is suitable for super-gaussian source signals. The Gaussian, Laplacian, and the generalized Gaussian density models are also used to form other score functions, however we should note that these nonlinearities can only be applied to symmetric source densities but not to the asymmetric distributions [15]. In [9] a Pearson system based activation function has been proposed for a wide class of densities. The Pearson system score function is in the form of $\mathbf{g}(\mathbf{y}) = \frac{-(\mathbf{y}-a)}{b_0 + b_1\mathbf{y} + b_2\mathbf{y}^2}$. The unknown variables (a, b_0, b_1, b_2) are easily obtained by the method of moments[9]. The Pearson system based nonlinearity is just an appropriate choice for source distributions which are close to normal distributions.

THE NEW ADAPTIVE NONLINEAR ACTIVATION FUNCTION

Since, there is no fixed nonlinearity able to satisfy all kinds of source distributions, we look for an adaptive activation function adapting itself according to different densities and being a suitable choice for all kinds of distributions. The optimal nonlinearity which is in the form of 8, satisfies the stability condition of the learning rules algorithms. Therefore, in order to obtain an optimal general adaptive score function we should try to make a close estimate to the true PDF .

For this purpose we use the FM based MAXENT technique as a powerful tool for density estimation. Since in BSS problems, after mixing the sources, the observed signals may be negative, we will first use a transform without any discontinuity that makes the random signal positive and then apply our method to estimate the PDF . For instance, if we use the exponential transform, the new nonlinearity of the activation function for each of the observed signals y_i, simply takes the form

$$g(y_i) = \sum_{j=0}^{j=M} \lambda_j \alpha_j \exp(\alpha_j y_i) - 1 \tag{9}$$

The new score function has a simple low order form and its coefficients will be obtained adaptively according to the observed signals fractional moments.

SIMULATIONS AND RESULTS

We have set up an experiment with four independent sources of exponential density family, mixing sources with Rayleigh(1), Gamma(1,1), Weibull(1.5,1.6) and Chi-Square (with 3 degrees of freedom) distribution models. The sources will be transformed into zero-mean variates and the mixing matrix will be generated randomly. In figure 2 we have shown the result of signal extraction using the FM MAXENT score function ($M = 2$). We have compared in figure 3 our new nonlinearity performance with common widely used nonlinearities in FASTICA package[10], pearson system based score

function [9] and Jade algorithm [11] and also the ordinary MAXENT method [12] with degree four . As a measure of quality performance, we utilize Amari's measure "inter-channel interference"(ICI) defined as [15]

$$J(P) = \frac{1}{Ns} \sum_{i=1}^{Ns} \frac{\sum_{k=1}^{Ns} |p_{ik}|^2}{\max_k |p_{ik}|^2} - 1 \tag{10}$$

where p_{ij} are the matrix $P = WA$ arrays and Ns is the number of sources. In figure 3 the ICI measure performance of various nonlinearities versus the number of samples are plotted. In each case, the performance has been averaged for 1000 realizations. In our simulations the number of iterations in the updating learning rule is set to 10 times. As it can be seen, the new adaptive activation function has a significant amount of performance improvement in contrast to the other well-known ones.

CONCLUSION

In this paper we have proposed a new adaptive nonlinear activation function . This new score function is based on the MAXENT PDF estimation by means of the optimal FM. The new simple form of the activation function performs better than the existing blind ICA approaches regardless of the source distribution models. We compared our experiment with the well-known BSS algorithms with various nonlinear score functions. The results show a promising performance using the new adaptive nonlinearity.

REFERENCES

1. H. Mathis and S. C. Douglas, "On the Existence of Universal Nonlinearities for Blind Source Separation", *IEEE Trans on Signal Processing*, vol. 50, No. 5, May 2002.
2. C. M. Bender, S. A. Orszag, "*Advanced Mathematical Methods for Scientists and Engineers*," McGraw–Hill, 1978.
3. E. T. Jaynes, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, Vol. 70, pp. 939-952, 1982.
4. H. Gzyl, P. N. Inveradi, A. Tagliani, M. Villasana, "Maxentropic solution of fractional moment problems," *Applied Mathematics and Computation*, 2005.
5. A. Tagliani,"On the proximity of distributions in terms of coinciding fractional moments," *Applied Mathematics and Computation*, 145, pp. 501-509, 2003.
6. P. N. Inveradi, A. Petri, G. Pontuale, A. Tagliani,"Stieltjes moment problem via fractional moments," *Applied Mathematics and Computation*, 166, pp. 664-677, 2005.
7. J. Cardoso, "Infomax and maximum likelihood for source separation", *IEEE Signal Processing Letters*,Vol. 4, Apr 1997.
8. H. Mathis and S. C. Douglas, "On optimal and universal nonlinearities for blind signal separation", *Proc. IEEE Int. Conf. Acoust. , Speech, Signal Process*, May 2001.
9. J. Karvanen and V. Koivunen, "Blind separation methods based on Pearson system and its extensions", *Signal Processing*, Vol. 84, 2002.
10. A. Hyvarinen, "*http://www.cis.hut.fi/projects/ica/fastica*"
11. J. Cardoso, "*http://sig.enst.fr/~cardoso*"
12. B. A. Bastami and H. Amindavar, "A New Adaptive General Non-Linear Activation Function for Blind Source Separation," *Accepted for presentation in the Workshop on Transform Based on Independent Component Analysis for Audio, Video and Hyperspectral Images Data Reduction and Coding, TBICA 2006, CNRS, France.*

13. A. Tagliani, Y. VelAasquez,"Inverse Laplace transform for heavy-tailed distributions," *Applied Mathematics and Computation*, 150, pp.337-345, 2004.
14. J. Cardoso, "On the stability of source separation algorithms", *Proc. IEEE Int. Workshop Neural Networks Signal Process*,Sept 1998.
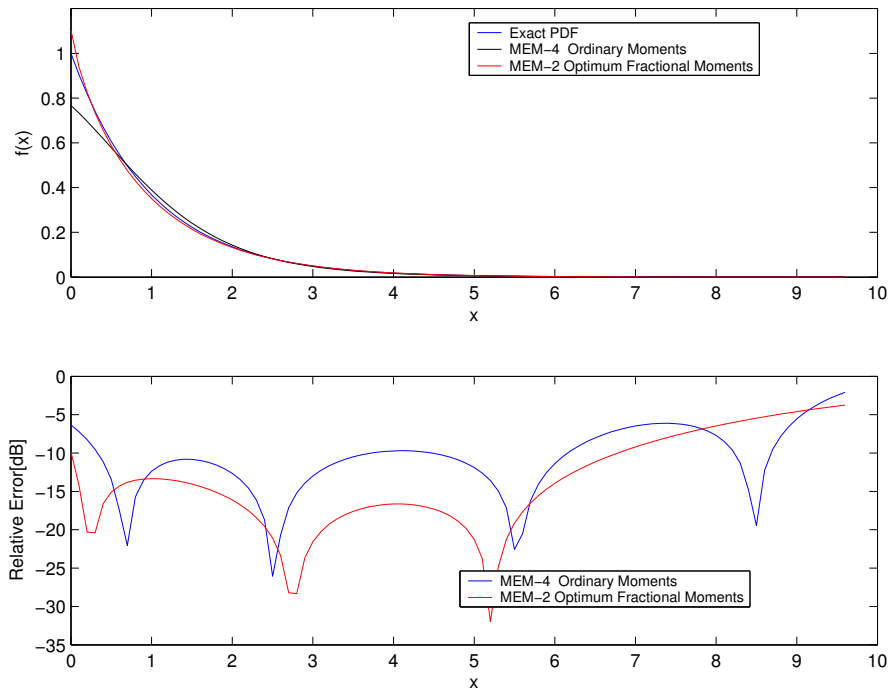15. A. Cichocki, S. Amari, "Adaptive blind signal and image processing ", *John Wiley and Sons*,2003.

FIGURE 1. Comparison of MAXENT density estimates for a Gamma distributed random variable($a = 1, b = 1$), using four integer moments, and two optimum FM.
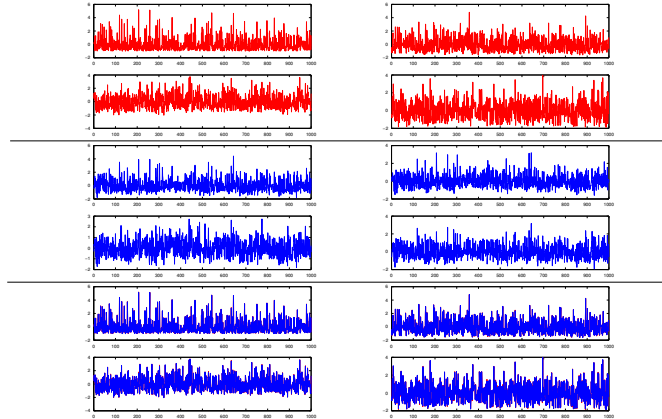
FIGURE 2. Signal extraction result using the MAXENT FM score function ($M = 2$), the sources are Rayleigh(1), Gamma(1,1), Weibull(1.5,1.6) and Chi-Square (with 3 degrees of freedom), distributed which are transformed into zero mean variates,the number of samples is 1000, the number of iterations in the updating learning rule is 10, the figures show the four normalized zero mean sources in the first row and the mixed observed signals in the second row. In the third row the separated and the source signals have been shown in a same plot.
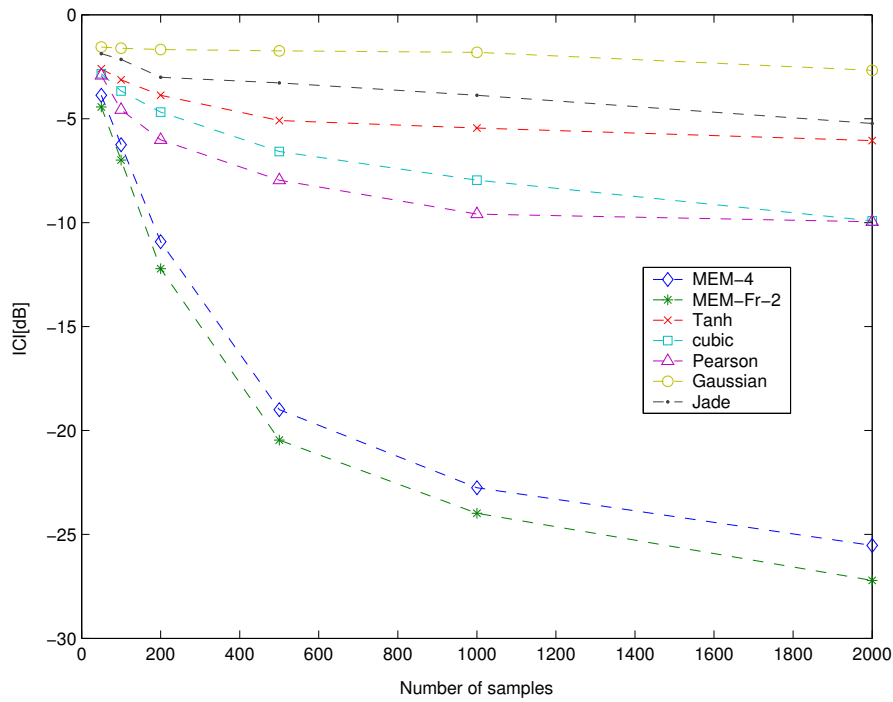

FIGURE 3. Comparison between the performances of different algorithms in BSS versus the number of samples, the mixed sources distributions models are Rayleigh(1), Gamma(1,1), Weibull(1.5,1.6) and Chi-Square (with 3 degrees of freedom), the number of iterations in the updating learning rule is 10.The Performances are averages over 1000 realizations.