Turk J Elec Eng & Comp Sci, Vol.20, No.3, 2012, © TÜBİTAK
doi:10.3906/elk-1004-1

Performance comparison of new nonparametric independent component analysis algorithm for different entropic indexes

Yusuf SEVİM, Ayten ATASOY*
Department of Electrical and Electronics Engineering, Faculty of Engineering, Karadeniz Technical University, 61080 Trabzon-TURKEY
e-mails: ysevim@ktu.edu.tr, ayten@ktu.edu.tr

Received: 01.04.2010

Abstract

Most independent component analysis (ICA) algorithms use mutual information (MI) measures based on Shannon entropy as a cost function, but Shannon entropy is not the only measure in the literature. In this paper, instead of Shannon entropy, Tsallis entropy is used and a novel ICA algorithm, which uses kernel density estimation (KDE) for estimation of source distributions, is proposed. KDE is directly evaluated from the original data samples, so it solves the important problem in ICA: how to choose nonlinear functions as the probability density function (pdf) estimation of the sources.

Key Words: *Independent component analysis (ICA), kernel density estimation (KDE), shannon entropy, tsallis entropy, mutual information (MI), nonparametric ICA (NpICA)*

1. Introduction

Independent component analysis (ICA) has attracted much attention recently for blindly separating mixtures of signals generated by independent sources. As a result, ICA techniques have rapidly growing applications in various subjects, such as image enhancement, biomedical signal processing, telecommunication, pattern recognition, and speech processing [1]. Several efficient ICA algorithms have been developed for applications in these areas [2-5]. Despite such wide application areas, most of the current ICA algorithms can separate mixed signals for some cases, but the performances of the algorithms are not so good. These performance weaknesses stem from 2 factors: the entropy estimation technique and the optimization of the entropy performance function [6].

After Shannon's famous work, there have been numerous generalizations of Shannon entropy, but only Tsallis entropy [7] has succeed in a natural generalization of the usual Boltzmann-Gibbs statistical mechanics

*Corresponding author: Department of Electrical and Electronics Engineering, Faculty of Engineering, Karadeniz Technical University, 61080 Trabzon-TURKEY

[8,9]. As a result of this, the Tsallis entropy measure is finding a place in many applications, including optimization [10] and electroencephalographic (EEG) signal analysis [11]. Indeed, it is a generalization of Shannon entropy and was inspired by the fractal behavior of many phenomena. Before Tsallis, Havrda and Charvat introduced similar forms of this entropy measure in 1967 [12]. Therefore, this entropy is subsequently called Havrda-Charvat-Tsallis (H-C-T) entropy.

Like other entropy measures, Shannon entropy can be retrieved from Tsallis entropy in the $q \rightarrow 1$ limit. In addition, Tsallis entropy has the remarkable property, not yet understood, that in the case of independent signal experiments, it does not satisfy the additivity property. Hence, Tsallis entropy is also termed nonextensive statistics. Another interesting property of Tsallis entropy is that it can present nice interpretations for long-range interactions and long-term memories, which cannot be explained systematically in the usual Boltzmann-Gibbs statistical mechanics [9].

To investigate Tsallis entropy for an ICA algorithm, in this paper, Tsallis entropy measure is utilized as an alternative to Shannon entropy with different order parameters (q). Unlike classical ICA algorithms that rely on a priori assumptions about the sources and mixing matrix for determining the unknown unmixing matrix, the new ICA algorithm uses the kernel density estimation (KDE) method to estimate the probability density function (pdf) of the reconstructed signals. Most ICA algorithms restrict the probability distribution of the sources to a particular shape or class via fixed or parametric estimates [7-12]. When the pdf of the source signals matches the restricted probability distribution of the sources, the ICA algorithms are capable of quickly and efficiently achieving the desired source separation. However, such algorithms can perform suboptimally or even fail to separate the desired source signals when the assumed statistical model is inaccurate [13]. As a solution for this problem, a more flexible model for pdfs of the source signals that consists of kernel-based techniques has been exploited [14].

In this paper, a new ICA algorithm is developed using the nonadditivity property of Tsallis entropy. A review of Tsallis entropy is given in Section 2. The actual algorithmic implementations of the new and the old techniques are given in Section 3. Simulations were conducted to demonstrate the performance comparisons obtained with the nonparametric ICA (NpICA) and the new ICA algorithm; these are given in section 4. Section 5 contains concluding remarks.

2. Tsallis entropy

The entropy of a discrete source is often obtained from its pdf, where $p = \{p_i\}$ is the probability of finding the system in each possible state i. Therefore, $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} p_i = 1$, where k is the total number of states. Using the pdf definition, Shannon entropy can be written as:

$$H = -\sum\nolimits_{i=1}^{k} p_i \log{(p_i)}. \tag{1}$$

Like Shannon entropy, the Tsallis entropy of random variable is defined from its pdf as:

$$S_q = \frac{1 - \sum_{i=1}^{k} (p_i)^q}{q-1}, \tag{2}$$

where the real number q is an entropic index that characterizes the degree of nonextensivity. Note that Tsallis entropy is concave for any $q \in R^+$ [15,16].

It can be noticed that, since in all cases $S_q \geq 0$, then $q < 1, q = 1$, and $q > 1$ respectively correspond to superadditivity, addivity, and subadditivity. The naming of these cases can be easily understood if it is noted that $p_i^q > p_i$ for $q < 1$ (the superadditive case), because the probability values are between 0 and 1. In this sense, Tsallis entropy presents us with a mean emphasizing the rare and frequent events.

Another general structure underlying the definition of Tsallis entropy is given in terms of deformed exponentials. In this framework, a q-deformed logarithm and q-expectation E_q are defined as follows.

$$\ln_q(x) = \frac{x^{1-q} - 1}{1-q} \tag{3}$$

$$E_q(x) = \frac{\int x p_x(x)^q dx}{\int p_{x'}(x')^q dx} \tag{4}$$

The q-exponential is defined as:

$$\exp_q(x) = (1 + (1-q)x)^{\frac{1}{1-q}}. \tag{5}$$

Eqs. (3) and (5) become the ordinary logarithm and exponential in the $q \rightarrow 1$ limit. Now Tsallis entropy can be rewritten in terms of the q-deformed logarithm as:

$$S_q = \sum\nolimits_{i=1}^{k} p_i \ln_q \left({}^{1}/_{p_i} \right). \tag{6}$$

A distinctive property of Tsallis entropy is that it is nonextensive; that is, the entropy of the joint density of 2 independent random variables is not equal to the sum of the entropies of random variables. Instead, it fulfills the following relation:

$$S_q(A,B) = S_q(A) + S_q(B) + (q-1)S_q(A)S_q(B), \tag{7}$$

where A and B are 2 independent random variables with no correlation. The parameter q is associated with the nonadditivity of the entropy values of the independent random variables A and B. When $q = 1$, the nonadditivity of Eq. (7) reverts to the standard additivity, such that Shannon entropy is considered to be an extensive entropy in contrast to Tsallis entropy.

3. ICA algorithm

Typically, ICA is a mathematical technique for recovering latent source signals from observed signal mixtures. Depending on the mixing process, ICA can be modeled as linear, nonlinear, and convolutive. In this work, the mixing process is considered as instantaneous linear mixing:

$$x = As, \tag{8}$$

where the sources $s = [s_1, ..., s_m]^T$ are mutually independent and A_{mxm} is an unknown mixing matrix. The aim is to find a matrix, W, only from observations, x, such that the output

$$y = Wx \tag{9}$$

is an estimate of the possible sources. Due to the fact that source signals are not known, estimating the demixing matrix W in the closed form cannot be possible. For this reason, different cost functions were proposed in the literature [4,17,18]. The solutions to ICA algorithms are found at the minimum or maximum of these cost functions. Minimization of these functions, possibly under some constraints on the solutions, is the subject of the optimization theory.

Using basic information theory equalities [19], mutual information (MI) can be written as a cost function for ICA algorithms. The reason for this is that MI is a measure of dependence between the random variables, and it is equal to zero if and only if the random variables are independent. Boscolo et al. used MI with the KDE method in their works on NpICA [3,20]. MI is defined between m (scalar) random variables y_i, $i = 1...m$, as follows.

$$I(y_1, y_2, ..., y_m) = \sum_{i=1}^{m} H(y_i) - H(y), \tag{10}$$

where $H(y_i)$ are the marginal entropies of the random variables and $H(y)$ is the joint entropy of the random variables. Minimization of the MI between the reconstructed signals can be written as:

$$\min_{W} \left\{ \sum_{i=1}^{m} H(y_i) - \log|\det W| - H(x) \right\}. \tag{11}$$

In Eq. (10), calculation of joint probability density is a burdensome task. Therefore, in Eq. (11), linear transformation is applied to the joint probability density; it can be seen that the entropy of the joint probability density of x is a constant with respect to W. As a result of this, the cost function is reduced to the following:

$$L(W) = \sum_{i=1}^{m} H(y_i) - \log|\det W|, \tag{12}$$

$$L(W) = \sum_{i=1}^{m} E[\log p_{y_i}(w_i x)] - \log|\det W|, \tag{13}$$

where w_i is the ith row of the matrix W. In order to evaluate Eq. (13), entropy definitions and a model for the distribution of the unknown signals are required. As a solution for this, we use KDE techniques, where the pdfs p_{y_i} are directly estimated from the data [21,22].

Another advantage of KDE techniques is that they allow direct evaluation of the cost function, and thus separating the optimization step from the step involving the reestimation of the score function, as in [23,24], is not necessary [3]. Using the KDE technique, the marginal distribution of a dataset is approximated as follows:

$$p_{y_i}(y_i) = \frac{1}{Mh} \sum_{m=1}^{M} \varphi\left(\frac{y_i - Y_{im}}{h}\right) \quad , \quad i = 1, ..., N, \tag{14}$$

where M is the size of the sample data, h is the kernel bandwidth $\left(h = 1.06M^{-1}/5\right)$, and φ is the Gaussian kernel:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2}/2. \tag{15}$$

The kernel centroids are equal to:

$$Y_{im} = w_i x^{(m)} = \sum_{n=1}^{N} w_{in} X_{nm}, \tag{16}$$

where $x^{(m)}$ is the mth column of the mixture matrix X. Replacing p_{y_i} in Eq. (13) with Eq. (14), the cost function can be written as:

$$L(W) = \sum_{i=1}^{N} E \log\left[\frac{1}{Mh}\sum_{m=1}^{M}\varphi\left(\frac{y_i - Y_{im}}{h}\right)\right] - \log|\det W|. \tag{17}$$

The overall optimization problem can be written as:

$$\min_{W}\left\{-\frac{1}{M}\sum_{i=1}^{N}\sum_{k=1}^{M}\log\left[\frac{1}{Mh}\sum_{m=1}^{M}\varphi\left(\frac{y_i - Y_{im}}{h}\right)\right] - \log|\det W|\right\}. \tag{18}$$

Optimizing the cost function on the unit sphere, this step must be complemented by projecting w on the unit sphere by dividing w by its norm.

$$w \leftarrow \frac{w}{\|w\|} \tag{19}$$

A new ICA algorithm using Tsallis entropy as the entropy measurement can be established by utilizing Kullback-Leibler divergence, D, between the random variable Y of n dimensions, and the product of its components, $\prod_{i=1}^{n} Y_i$, is derived as in [16,25].

$$D_q\left(Y \middle| \prod_{i=1}^{N} Y_i\right) = -S_q(Y) + E_q\left[\ln_q \prod_{i=1}^{N} p_{Y_i}(Y_i)\right], \tag{20}$$

where N is the number of the components of independent sources, and $\ln_q$ and E_q are defined in Eqs. (3) and (4), respectively. For 2 signals, the cost function of the new ICA algorithm using the KDE technique and Eq. (20) can be obtained as follows.

$$L(W) = \sum_{i=1}^{2} E_q \ln_q\left[\frac{1}{Mh}\sum_{m=1}^{M}\varphi\left(\frac{y_i - Y_{im}}{h}\right)\right] + (q-1)\, E_q \ln_q\left[\frac{1}{Mh}\sum_{m=1}^{M}\varphi\left(\frac{y_1 - Y_{1m}}{h}\right)\right] E_q \ln_q\left[\frac{1}{Mh}\sum_{m=1}^{M}\varphi\left(\frac{y_2 - Y_{2m}}{h}\right)\right]$$

$$-E_q \ln_q\left[\frac{1}{Mh^2}\sum_{m=1}^{M}\prod_{d=1}^{2}\varphi\left(\frac{y_d - Y_{dm}}{h}\right)\right] \tag{21}$$

when Eq. (21) is compared with Eq. (17), it can be seen that the calculation complexity of the new ICA algorithm increases in cases of $m \geq 3$. The other difficulty of the new ICA algorithm is due to the intractable computational cost of estimation for the joint density function of reconstructed signals. As a partial solution to this problem, linear transformation can be applied to the Tsallis entropy of the joint density function of the reconstructed signals; that is:

$$E_q \ln_q\left[\frac{\frac{1}{Mh^2}\sum_{m=1}^{M}\prod_{d=1}^{2}\varphi\left(\frac{x_d - X_{dm}}{h}\right)}{|\det W|}\right], \tag{22}$$

where x is the observed signals. Eq. (22) calculates the Tsallis entropy of y with respect to x and $|\det W|$. It is obvious that this equation cannot be eliminated in the optimization stage. Thus, it is necessary to estimate the joint density of the observed signals only once, and then using this estimation and Eq. (22), the Tsallis entropy estimation of the joint density of the reconstructed signals can be estimated for every step of the optimization stage. For simplicity, the first derivatives of the cost function (m =2) with respect to w_{ij} are computed as:

$$\frac{\partial L(W)}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} S_q(Y_i) + (q-1) S_q(Y_{3-i}) \frac{\partial}{\partial w_{ij}} S_q(Y_i) - \frac{\partial}{\partial w_{ij}} S_q(Y), \tag{23}$$

where $S_q(Y_i)$ is the Tsallis entropy of the marginal distributions (i, j= 1,2) and $S_q(Y)$ is the Tsallis entropy of the joint distribution. $\frac{\partial}{\partial w_{ij}} S_q(Y_i)$ can be calculated as follows.

$$\frac{\partial}{\partial w_{ij}} S_q(Y_i) = \frac{1}{q-1} \frac{\frac{\partial}{\partial w_{ij}} \left[\sum_{t=1}^{N} \sum_{k=1}^{M} \frac{1}{Mh} \varphi \left(\frac{w_t^{(k)} \left(x_t^{(k)} - x^{(m)} \right)}{h} \right) \right]^q}{\sum_{t=1}^{N} \sum_{k=1}^{M} \frac{1}{Mh} \varphi \left(\frac{w_t^{(k)} \left(x_t^{(k)} - x^{(m)} \right)}{h} \right)} \tag{24}$$

$$\frac{\partial}{\partial w_{ij}} S_q(Y_i) = \frac{1}{q-1} \frac{-\frac{q}{(Mh)^q h} \sum_{t=1}^{N} \sum_{k=1}^{M} x^{(m)} \left(y_t - w_t x^{(m)} \right) \varphi \left(\frac{w_t^{(k)} \left(x_t^{(k)} - x^{(m)} \right)}{h} \right)}{\sum_{t=1}^{N} \sum_{k=1}^{M} \frac{1}{Mh} \varphi \left(\frac{w_t^{(k)} \left(x_t^{(k)} - x^{(m)} \right)}{h} \right)} \tag{25}$$

It is easy to extend to the case of any $m \in N$, but the calculation becomes very complicated.

The final step for ICA algorithms is to minimize the cost functions. In this respect, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) [26] algorithm, derived from Newton's method, is utilized for the new ICA algorithm and Boscolo's ICA algorithm in optimization. In the BFGS algorithm, the Hessian matrix of second derivates of the cost function to be minimized does not need to be computed at any stage. The Hessian matrix is updated by analyzing successive gradient vectors, instead. The BFGS algorithm can be summarized as follows.

Step 1. Given $x_1 \in R^n$, $B_1 \in R^{nxn}$ positive definite, $f(x)$ is a function,
compute $g_1 = \nabla f(x_1)$. If $g_1 = 0$, stop; otherwise, set $k = 1$.

Step 2. Set $d_k = -B_k^{-1} g_k$.

Step 3. Carry out a line search along d_k, getting $\alpha_k > 0$,
$x_{k+1} = x_k + \alpha_k d_k$, and $g_{k+1} = \nabla f(x_{k+1})$.
If $g_{k+1} = 0$ or $(f(x_{k+1}) - f(x_k)) < 10^{-10}$, stop.

Step 4. Set

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k},$$

where

$$s_k = \alpha_k d_k,$$

$$y_k = g_{k+1} - g_k.$$

Step 5. $k := k + 1$; go to Step 2.

Using Eqs. (23) and (25) in the BFGS algorithm, the new ICA algorithm can be obtained for m= 2.

A well-known problem in ICA algorithms is that, although the global minimum of Eq. (11) is known to yield the desired source separation, there is no proof to show that such a function has no local minimum. Therefore, Comon [27] proved that convergence to any local minimum rather than global minimum would result in a failure to separate the source signals. Rattray and Basalyga [28] and Rattray [29] showed that this issue can be overlooked in ICA algorithms, and Amari [17] showed that ICA algorithms can separate source signals for arbitrary initial guess. In light of these works, convergence conditions were not investigated in this study.

4. Performance evaluation

Experiments were performed using 5 synthetic data sets for different pdfs. Whereas 4 of them had super-Gaussian and sub-Gaussian pdfs, the last 1 had a Gaussian pdf. To validate the new ICA algorithm using the nonparametric method as a pdf estimation method, synthetic datasets were adopted as input mixtures.

In order to produce the test data, 1000 synthetic datasets were linearly mixed by 1000 randomly generated mixing matrixes with sample sizes ranging between 512 and 4608. One of the source signals and one of the synthetic datasets are shown in Figures 1 and 2, respectively.

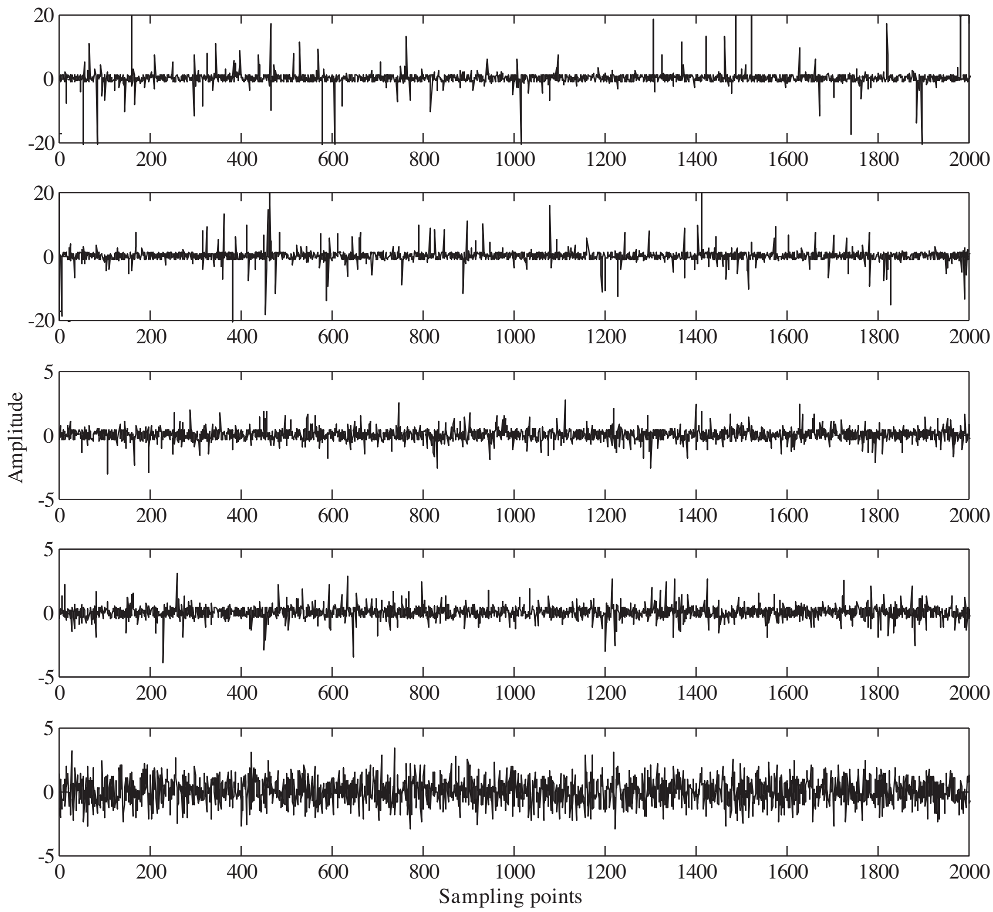

Figure 1. Source signals.

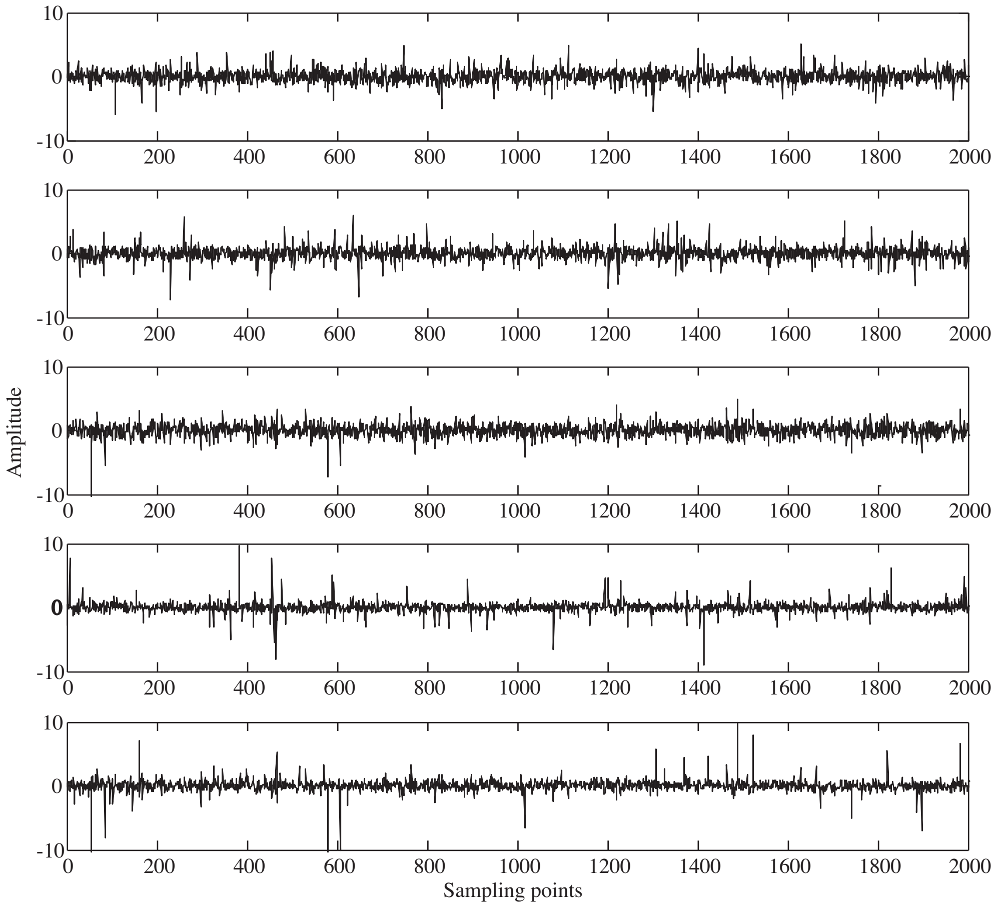

Figure 2. Synthetic signals.

The NpICA algorithm was compared with the new ICA algorithm for different entropic indexes ($q = 1.1, ..., 1.5$) on these datasets. These algorithms were quantitatively compared using only one measure, which was the median signal-to-noise ratio (SNR) of the separated signals. The equation describing the SNR is:

$$SNR = 10\log_{10}\left(\frac{E\left[s(t)^2\right]}{E\left[n(t)^2\right]}\right), \tag{26}$$

where $E(.)$is the mean of the arguments, $s(t)$is the desired source signal, and $n(t) = s(t) - \hat{s}(t)$ is the noise, indicating an undesired signal. Here $\hat{s}(t)$ is the estimated source signals, and $s(t)$ and $\hat{s}(t)$ are at the same energy.

The results of these experiments are shown in Figure 3, and it can be seen clearly that raising the number of samples increases performance gains obtained with both nonparametric ICA algorithms. Nevertheless, for the new ICA algorithm, increasing the q entropic index decreases the median SNR. If median SNR levels are below 8-10 dB, this indicates that the system has failed to obtain the desired source separation.

For the super-Gaussian sources (Source 1 and 2), the median SNR is higher than those for the other sources. It can be seen in Figure 3 that the Gaussian source, Source 5, has the lowest gains among the sources. Although for $q = 1.1$ and 1.2, the new ICA algorithm appears to match the performance of Boscolo's algorithm, both algorithms appear to be capable of separating the sources even when the sample size is very small (e.g. 512 samples).

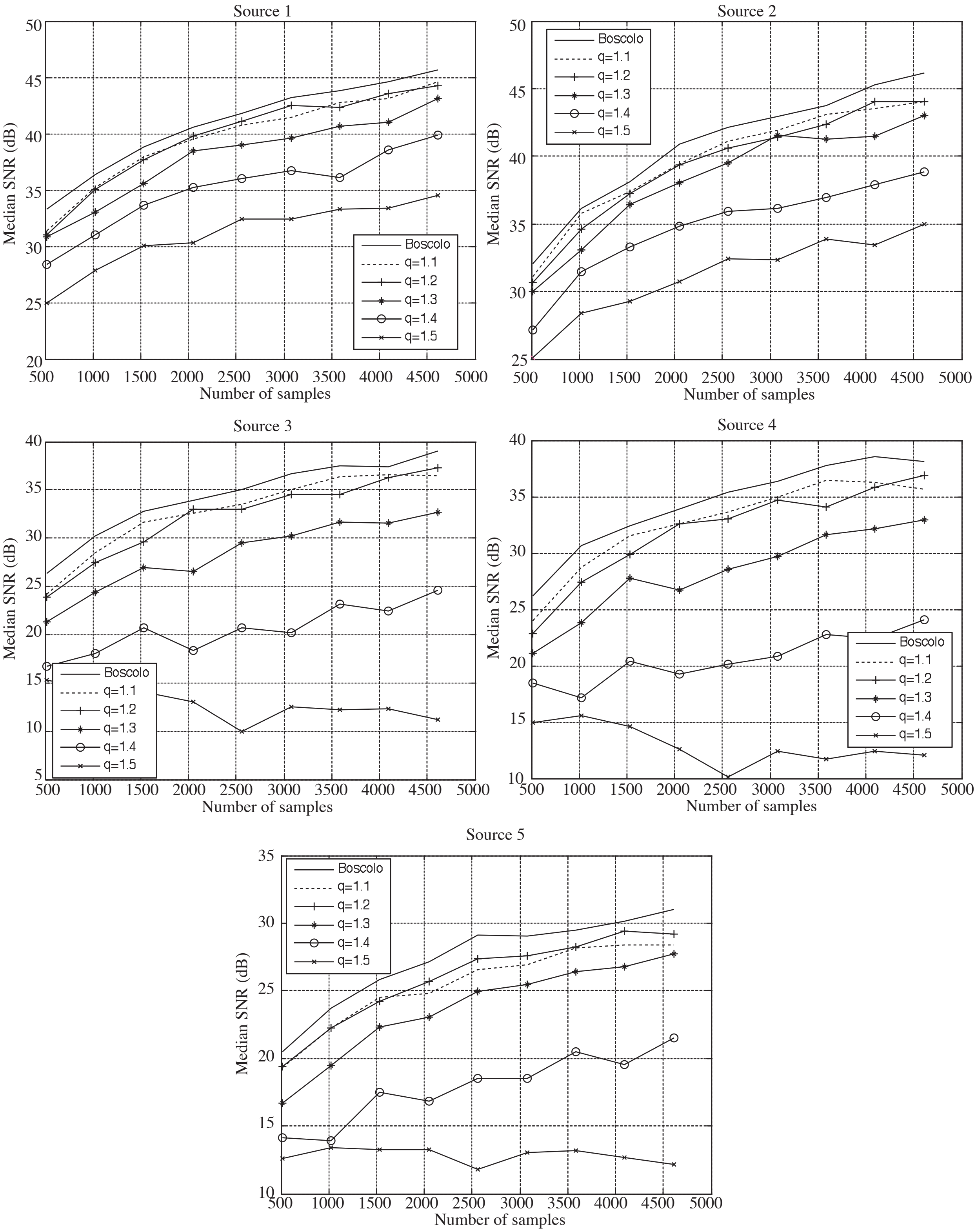

Figure 3. The results of separation of the 5 different sources are shown for nonparametric ICA algorithms (averaged over 1000 simulations).

5. Conclusions

The properties of the proposed method in this paper were investigated for large-scale problems and different entropy indexes. For this purpose, the results of the proposed method were compared with the results of the NpICA algorithm for synthetic datasets. It can be noticed that neither algorithm needs the distributions of the source signals and parameters or nonlinearity for the contrast functions.

The results of the synthetic datasets demonstrate that although the proposed algorithm is capable of separating signals, noticeable performance improvement cannot be seen when the performance of the proposed algorithm is compared with that of the NpICA algorithm ($q > 1.1$). Nevertheless, the performance of the proposed method comes closer to the performance of the NpICA algorithm when the entropy index reaches a value of 1. From the other point of view, the cost function and calculation of the entropy of the proposed algorithm are different from the NpICA algorithm. A lower SNR value is therefore very natural for different entropy indexes. On the other hand, the proposed algorithm has a lower step size (1 or 2, as averaged over 1000 simulations).

The other important result is that if the source signals have super-Gaussian pdfs, the performances of both algorithms are higher than performances of other source signals with sub-Gaussian or Gaussian pdfs. Another parameter that increases the performances of the algorithms is the sample size. It can be seen clearly from the Figures that increasing the sample size improves the performances of the algorithms.

Acknowledgments

This study was supported by Karadeniz Technical University Scientific Research Projects Unit, Project No: 2007.112.04.2.

References

[1] A. Cichocki, S. Amari, Adaptive Blind Signal and Adaptive Blind Signal and Image Processing, Chichester, John Wiley & Sons, 2002.

[2] A. Hyvarinen, "Survey on independent component analysis", Neural Computating Surveys, Vol. 2, pp. 94-128, 1999.

[3] R. Boscolo, H. Pan, V.P. Roychowdhury, "Independent component analysis based on non-parametric density estimation", IEEE Transactions on Neural Networks, Vol. 15, pp. 55-65, 2004.

[4] A. Bell, T. Sejnowski, "An information maximization approach to blind source separation and blind deconvolution", Neural Computing, Vol. 7, pp. 1129-1159, 1995.

[5] J.F. Cardoso, "High-order contrasts for independent component analysis", Neural Computation, Vol. 11, pp. 157-192, 1999.

[6] D.J. Krusienski, W.K. Jenkins, "Nonparametric density estimation based independent component analysis via particle swarm optimization", ICASSP 2005, Vol. 4, pp. 357-360, 2005.

[7] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics", Journal of Statistical Physics, Vol. 52, pp. 479-487, 1988.

[8] C. Tsallis, Nonextensive Statistical Mechanics and Its Applications, edited by S. Abe and Y. Okamoto, Heidelberg, Springer Verlag, 2001.

[9] S. Abe, "Tsallis' nonextensive statistical mechanics", Mathematical Science, Vol. 38, pp. 71-80, 2000.

[10] A. Fachat, K.H. Hoffmann, A. Franz, "Simulated annealing with threshold accepting for Tsallis statistics", Computer Physics Communications, Vol. 132, pp. 232-240, 2000.

[11] A. Capurro, L. Diambra, D. Lorenzo, O. Macadar, M.T. Martin, C.M. Mostaccio, A. Plastino, E. Rofman, M.E. Torres, J. Velutti, "Tsallis entropy and cortical dynamics: the analysis of EEG signals", Physica A, Vol. 257, pp. 149-155, 1998.

[12] J.H. Havrda, F. Charvat, "Quantification methods of classification process: concepts of structural α entropy", Kybernetika, Vol. 3, pp. 30-35, 1967.

[13] J.F. Cardoso, "Blind signal separation: statistical principles", Proc. IEEE Special Issue on Blind Identification and Estimation, Vol. 9, pp. 2009-2025, 1998.

[14] F.R. Bach, M.I. Jordan, "Kernel independent component analysis", Journal of Machine Learning Research, Vol. 3, pp. 1-48, 2002.

[15] A.K. Rajagopal, S. Abe, "Implications of form invariance to the structure of nonextensive entropies", Physical Review Letters, Vol. 83, pp. 1711-1714, 1999.

[16] H. Suyari, "Nonextensive entropies derived from form invariance of pseudoadditivity", Physical Review E, Vol. 65, 2002. DOI: 10.1103/PhysRevE.65.066118.

[17] S. Amari, A. Cichocki, H.H. Yang, A New Learning Algorithm for Blind Signal Separation, Advances in Neural Information Processing Systems, Vol. 8, Cambridge, MIT Press, 1996, pp. 757-763.

[18] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis", IEEE Transactions on Neural Networks, Vol. 10, pp. 626-634, 1999.

[19] T.M. Cover, J.A. Thomas, Elements of Information Theory, New York, John Wiley & Sons, 1991.

[20] R. Boscolo, H. Pan, V.P. Roychowdhury, "Non-parametric ICA", Proceedings of the Third International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2001), San Diego, California, 2001.

[21] M.C. Jones, The Projection Pursuit Algorithm for Exploratory Data Analysis, PhD thesis, School of Mathematics, University of Bath, 1983.

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

[23] N. Vlassis, Y. Motomura, "Efficient source adaptivity in independent component analysis", IEEE Transactions on Neural Networks, Vol. 12, pp. 559-566, 2001.

[24] J. Karvanen, J. Eriksson, V. Koivunen, "Pearson system based method for blind separation", Proceedings of Second International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, pp. 585-590, 2000.

[25] H. Suyari, M. Dakemoto, "On independent component analysis using Tsallis mutual Entropy", International Symposium on Nonlinear Theory and its Application, Xi'an, China, 2002.

[26] E.K.P. Chong, S.H. Zak, An Introduction To Optimization, New York, John Wiley & Sons, 2001.

[27] P. Comon, "Independent component analysis, a new concept?", Signal Processing, Vol. 36, pp. 287-314, 1994.

[28] M. Rattray, G. Basalyga, "Scaling laws and local minima in Hebbian ICA", in Advances in Neural Information Processing Systems 14 (T.G. Dietterich, S. Becker, Z. Ghahramani, Editors), Cambridge, MIT Press, pp. 495-501, 2001.

[29] M. Rattray, "Stochastic trapping in a solvable model of on-line independent component analysis", Neural Computation, Vol. 14, pp. 421-435, 2002.