Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

NEUROCOMPUTING

ELSEVIER

Neurocomputing 68 (2005) 267–272

www.elsevier.com/locate/neucom

Letters

An alternative switching criterion for independent component analysis (ICA)☆

Dengpan Gao, Jinwen Ma*, Qiansheng Cheng

Department of Information Science, School of Mathematical Sciences and LMAM, Peking University, Beijing 100871, PR China

Received 2 November 2004; received in revised form 17 April 2005; accepted 18 April 2005
Available online 13 June 2005
Communicated by R.W. Newcomb

Abstract

In solving the problem of noiseless independent component analysis (ICA) in which sources of super- and sub-Gaussian coexist in an unknown manner, one can be lead to a feasible solution using the natural gradient learning algorithm with a kind of switching criterion for the model probability distribution densities to be selected as super- or sub-Gaussians appropriately during the iterations. In this letter, an alternative switching criterion is proposed for the natural gradient learning algorithm to solve the noiseless ICA problem with both super- and sub-Gaussian sources. It is demonstrated by the experiments that this alternative switching criterion works well on the noiseless ICA problem with both super- and sub-Gaussian sources.

Keywords: Independent component analysis; Switching criterion; Mixed signals; Sub- and super-Gaussian; Kurtosis

☆This work was supported by the Natural Science Foundation of China for Project 60471054.
*Corresponding author. Tel.: +86 10 62758101; fax: +86 10 62751801.
E-mail address: jwma@math.pku.edu.cn (J. Ma).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2005.04.003

1. Introduction

The noisy or general independent component analysis (ICA) [2] aims at blindly separating the independent sources $\mathbf{s}$ from the observations $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{v}$ via

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \quad \mathbf{x}, \mathbf{v} \in \mathbb{R}^m, \quad \mathbf{y} \in \mathbb{R}^n, \quad \mathbf{W} \in \mathbb{R}^{n \times m}, \tag{1}$$

where $\mathbf{A}$ is a mixing matrix, $\mathbf{v}$ is a vector of uncorrelated noise terms, and $\mathbf{W}$ is the so-called de-mixing matrix to be estimated. Because the estimation of a de-mixing matrix $\mathbf{W}$

feasible solution of the ICA problem. It was demonstrated in simulation experiments that the natural gradient learning algorithm with this switching criterion can solve the ICA problem successfully in the cases of both super- and sub-Gaussian sources.

In this letter, we propose an alternative switching criterion directly from the point of view of the one-bit-matching principle, which is demonstrated for separating well the sources of both super- and sub-Gaussians from their linearly mixed signals.

2. The alternative switching criterion

The natural gradient learning algorithm with a switching criterion can be expressed as follows (refer to [9] for details):

$$\Delta \mathbf{W} = \eta[\mathbf{I} - \mathbf{K}\tanh(y)y^{\mathrm{T}} - yy^{\mathrm{T}}]\mathbf{W}, \tag{3}$$

where $\eta > 0$ is the learning rate, $\mathbf{K} = \mathrm{diag}[k_1, \ldots, k_n]$ is considered as a switching criterion represented by an n-order diagonal matrix with $k_i = 1$ and -1 for the super- and sub-Gaussian of model pdf $p_i(y_i)$, respectively, and $\tanh(y) = [\tanh(y_1), \tanh(y_2), \ldots, \tanh(y_n)]^{\mathrm{T}}$ with $\tanh(y_i)$ being the hyperbolic tangent function of y_i, and $m = n$. The model pdfs of sub- and super-Gaussians are actually selected as

$$p_{\mathrm{sub}}(u) = \tfrac{1}{2}[p_{N(1,1)}(u) + p_{N(-1,1)}(u)], \quad p_{\mathrm{super}}(u) \propto p_{N(0,1)}(u)\mathrm{sech}^2(u), \tag{4}$$

respectively, where $p_{N(\mu,\sigma^2)}(u)$ is the Gaussian or normal density with mean μ and variance σ^2, and $\mathrm{sech}(u)$ is the hyperbolic secant function. The switching criterion proposed by Lee et al. [9] is given by

$$k_i = \mathrm{sign}(E\{\mathrm{sech}^2(y_i)\}E\{y_i^2\} - E\{[\tanh(y_i)]y_i\}), \quad i = 1, 2, \ldots, n, \tag{5}$$

which is just a sufficient condition for the asymptotic stability of the solution of the ICA problem obtained by Cardoso and Laheld [4].

We now believe that the one-bit-matching conjecture on the natural gradient learning algorithm is true and construct a switching criterion from the point of view of the one-bit-matching condition alternatively. When the one-bit-matching condition is satisfied, the natural gradient learning algorithm will be finally stable at a feasible solution. i.e., with a certain permutation, the sign of kurtosis of model pdf $p_i(y_i)$ will be equal to the sign of kurtosis of the source pdf of s_i or equivalently the $p(y_i, \mathbf{W})$ of the resulting output y_i since $y_i = \lambda_i s_i$ with $\lambda_i \neq 0$. Thus, we have a sufficient condition on the stable and feasible solution of the ICA as follows:

$$\mathrm{sign}(\kappa(p_i(y_i))) = \mathrm{sign}(\kappa(p(y_i, \mathbf{W}))) = \mathrm{sign}(E\{y_i^4\} - 3E^2\{y_i^2\}), \quad i = 1, 2, \ldots, n, \tag{6}$$

where $\kappa(p_i(y_i))$ is the kurtosis of model pdf $p_i(y_i)$, while $\kappa(p(y_i, \mathbf{W})) = E\{y_i^4\} - 3E^2\{y_i^2\}$ is the kurtosis of $p(y_i, \mathbf{W})$ or the output signal y_i. So, we can design the sign of kurtosis of model pdf $p(y_i)$ as the sign of kurtosis of output y_i online during the iterations of the natural gradient learning algorithm so that the algorithm will tend to a feasible solution. Hence, we have the following switching criterion of k_i in

Eq. (3) as follows:

$$k_i = \mathrm{sign}(E\{y_i^4\} - 3E^2\{y_i^2\}), \quad i = 1, 2, \ldots, n. \tag{7}$$

Since $E(y_i^2), E(y_i^4)$ can be easily estimated from a sample data set of $\mathbf{x}$ via the relation $\mathbf{y} = \mathbf{W}\mathbf{x}$ in the noiseless case, this switching criterion or rule can be easily implemented by the estimation of the kurtosis of y_i during the iteration of the natural gradient learning algorithm. In fact, this kind of switching criterion was already suggested in the learning of the deflationary exploratory projection pursuit network [7].

3. Simulation results

We conducted the experiments on the noiseless ICA problem of five independent sources in which there are three super-Gaussian sources generated from the exponential distribution $E(0.5)$, the Gamma distribution $\gamma(1,4)$, and the Chi-square distribution $\chi^2(6)$, respectively, and two sub-Gaussian sources generated from the β distribution $\beta(2,2)$ and the Uniform distribution $U([0,1])$, respectively. From each distribution, 100 000 i.i.d. samples were generated to form a source. The linearly mixed signals were then generated from the five source signals in parallel via the following mixing matrix:

$$\mathbf{A} = \begin{bmatrix} 0.3987 & 0.1283 & 0.0814 & 0.7530 & 0.1213 \\ 0.1991 & 0.2616 & 0.1248 & 0.5767 & 0.7344 \\ 0.7607 & 0.4484 & 0.3950 & 0.7119 & 0.2145 \\ 0.4609 & 0.0275 & 0.0745 & 0.5171 & 0.0674 \\ 0.8369 & 0.5794 & 0.7752 & 0.4330 & 0.2373 \end{bmatrix}. \tag{8}$$

The learning rate was selected as $\eta = 0.001$ and the natural gradient learning algorithm operated in the adaptive mode and was stopped when all the 100 000 data points of the mixed signals had been passed only once through the updating or learning rule Eq. (3) with the switching criterion.

The result of the natural gradient learning algorithm with the alternative switching criterion Eq. (7) is given in Eq. (9), while the result of the natural gradient learning algorithm with the original switching criterion Eq. (5) is given in Eq. (10). As a feasible solution of the ICA problem, the obtained $\mathbf{W}$ will make $\mathbf{WA} = \mathbf{\Lambda P}$ be satisfied or approximately satisfied to a certain extent, where $\Lambda P = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]$ with each $\lambda_i \neq 0$, and P is a permutation matrix,

$$\mathbf{WA} = \begin{bmatrix} 0.0142 & 0.0034 & 1.4198 & 0.0380 & 0.0146 \\ 0.6805 & 0.0313 & -0.0020 & 0.0243 & -0.0008 \\ 0.0079 & 0.7439 & 0.0132 & -0.0147 & -0.0322 \\ 0.0037 & -0.0051 & -0.0027 & -0.0430 & 0.8376 \\ 0.0052 & -0.0041 & 0.0309 & 1.4245 & 0.0040 \end{bmatrix}, \tag{9}$$

$$\mathbf{WA} = \begin{bmatrix} 0.0141 & 0.0048 & 1.4198 & 0.0379 & 0.0145 \\ 0.6805 & 0.0312 & -0.0019 & 0.0243 & -0.0008 \\ 0.0080 & 0.7439 & 0.0125 & -0.0159 & -0.0315 \\ 0.0037 & -0.0058 & -0.0027 & -0.0431 & 0.8376 \\ 0.0053 & -0.0020 & 0.0311 & 1.4244 & 0.0042 \end{bmatrix} \tag{10}$$

From Eq. (9), we can observe that the natural gradient learning algorithm with the alternative switching criterion can really solve the noiseless ICA problem with both super- and sub-Gaussian sources since these degenerated elements in **WA** are rather small. In comparison with the results listed in Eq. (10), we can also find that the two switching criteria lead to almost the same solution. However, comparing Eq. (5) with Eq. (7), we can observe that $\text{sech}(y_i)$ and $\tanh(y_i)$ lead to certain additional computation in the original switching criterion and thus the computation cost of the alternative switching criterion is considerably less than that of the original one, which was indeed demonstrated by the experiments with the fact that the running time of the natural gradient learning algorithm with the alternative switching criterion is only about 0.5702 that of the natural gradient learning algorithm with the original switching criterion.

Furthermore, we conducted many other experiments on the different noiseless ICA problems with both super- and sub-Gaussian sources in which the natural gradient learning algorithm with the alternative switching criterion always arrives at a satisfactory result as above. Moreover, it was even found that as the number of sources increases, the alternative switching criterion gets a better solution of the ICA problem than the original switching criterion does.

4. Conclusions

An alternative switching criterion for the natural gradient learning algorithm to solve the noiseless ICA problem with both super- and sub-Gaussian sources is constructed with the help of the one-bit-matching conjecture. The experiments show that this alternative switching criterion is as good as the original one, but the computation cost is reduced considerably.

References

[1] S.I. Amari, A. Cichocki, H. Yang, A new learning algorithm for blind separation of sources, Adv. Neural Inf. Process. 8 (1995) 757–763.

[2] S.I. Amari, A. Cichocki, H.H. Yang, Blind signal separation and extracting: neural and information theoretical approaches, in: S. Haykin (Ed.), Unsupervised Adaptive, Filtering Blind Source Separation, vol. I, Wiley, New York, 2000, pp. 63–138.

[3] A. Bell, T. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (6) (1995) 1129–1159.

[4] J.F. Cardoso, B. Laheld, Equivalent adaptive source separation, IEEE Trans. Signal Process. 44 (12) (1996) 3017–3030.

[5] A. Cichocki, W. Kasprzak, S.I. Amari, Adaptive approach to blind source separation with cancellation of additive and convolution noise, in: Proceedings of the Third International Conference on Signal Processing (ICSP'96), vol. I, Beijing, China, 1996, pp. 412–415.
[6] P. Comon, Independent component analysis—a new concept?, Signal Process. 36 (3) (1994) 287–314.
[7] M. Girolami, C. Fyfe, Extraction of independent signal sources using a deflationary exporatory projection pursuit network with lateral inhibition, IEEE Proc. Vision Image Signal Process. 144 (5) (1997) 299–306.
[8] T.W. Lee, Independent Component Analysis: Theory and Applications, Kluwer Academic Publishers, Dordrecht, 1998.
[9] T.W. Lee, M. Girolami, T.J. Sejnowski, Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and supergaussian sources, Neural Comput. 11 (2) (1999) 417–441.
[10] Z.Y. Liu, K.C. Chiu, L. Xu, One-bit-matching conjecture for independent component analysis, Neural Comput. 16 (2) (2004) 383–399.
[11] J. Ma, Z. Liu, L. Xu, A further result on the ICA one-bit-matching conjecture, Neural Comput. 17 (2) (2005) 331–334.
[12] L. Xu, C.C. Cheung, S.I. Amari, Learned parametric mixture based ica algorithm, Neurocomputing 22 (1–3) (1998) 69–80.