A Further Result on the ICA One-Bit-Matching Conjecture

Jinwen Ma
jwma@math.pku.edu.cn
Department of Computer Science and Engineering, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, and School of Mathematical Sciences and LMAM, Peking University, Beijing, 100871, China

Zhiyong Liu
zyliu@cuhk.edu.hk
Lei Xu
lxu@cse.cuhk.edu.hk
Department of Computer Science and Engineering, Chinese University of Hong Kong, Shatin, N.T., Hong Kong

The one-bit-matching conjecture for independent component analysis (ICA) has been widely believed in the ICA community. Theoretically, it has been proved that under the assumption of zero skewness for the model probability density functions, the global maximum of a cost function derived from the typical objective function on the ICA problem with the one-bit-matching condition corresponds to a feasible solution of the ICA problem. In this note, we further prove that all the local maximums of the cost function correspond to the feasible solutions of the ICA problem in the two-source case under the same assumption. That is, as long as the one-bit-matching condition is satisfied, the two-source ICA problem can be successfully solved using any local descent algorithm of the typical objective function with the assumption of zero skewness for all the model probability density functions.

The so-called one-bit-matching conjecture, which states that "all the sources can be separated as long as there is a one-to-one same-sign-correspondence between the kurtosis signs of all source probability density functions and the kurtosis signs of all model probability density functions," was summarized in Xu, Cheung, and Amari (1998) and formally proved in Liu, Chiu, and Xu (2004) recently under the assumption of zero skewness for the model probability density functions. The one-bit-matching condition guarantees a successful solution of the ICA problem by globally maximizing the following cost function:

$$J(\mathbf{R}) = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^4 \nu_j^s k_i^m, \tag{1}$$

where the orthonormal matrix $\mathbf{R} = (r_{ij})_{n\times n} = \mathbf{WA}$ is to be estimated instead of $\mathbf{W}$ since $\mathbf{A}$, i. e., the mixing matrix, is a constant one,[1] ν_j^s is the kurtosis of the source s_j, and k_i^m is a constant with the same sign as the kurtosis ν_i^m of the model probability density $p_i(y_i)$. As proved by Liu et al. (2004), under the assumption of zero skewness for the model probability density functions, this cost function is equivalent to the typical objective function of the ICA problem used by several researchers, such as Bell and Sejnowski (1995), Amari, Cichocki, and Yang (1996), and Cardoso (1999).

However, in practice, typical gradient algorithms (e.g., Amari et al., 1996; Welling & Weber, 2001) cannot guarantee global optimization. Here we further analyze the local maxima of the cost function $J(\mathbf{R})$ in the case of two sources. For convenience, $J(\mathbf{R})$ can be rewritten as $\sum_{i=1}^n \sum_{j=1}^n r_{ij}^4 k_{ij}$, where $k_{ij} = \nu_j^s k_i^m$ is the element of matrix $\mathbf{K} = (k_{ij})_{n\times n}$.

Theorem. *Assume that the one-bit-matching condition is satisfied for $n = 2$, the local maxima of $J(\mathbf{R})$ are only the permutation matrices up to sign indeterminacy.*

Proof. For the second-order orthonormal matrix $\mathbf{R}$, there exist two disjoint fields of $\mathbf{R}$ that can be expressed by

$$\mathbf{R}_1 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

and

$$\mathbf{R}_2 = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix},$$

respectively, with one variable $\theta \in [0, 2\pi)$. Actually, each $\mathbf{R}_1$ denotes a rotation transformation in $\mathbb{R}^2$ (i.e., the two-dimensional real Euclidean space), while each $\mathbf{R}_2$ denotes a reflection transformation in $\mathbb{R}^2$.

Since $J(\mathbf{R}_1) = J(\mathbf{R}_2)$, for simplicity we consider just $\mathbf{R}_1$. By differentiating $J(\mathbf{R}_1)$ with respect to θ, we have

$$\frac{dJ(\mathbf{R}_1)}{d\theta} = 4\cos\theta\sin\theta[-(k_{11}+k_{22})\cos^2\theta + (k_{12}+k_{21})\sin^2\theta], \tag{2}$$

$$\frac{d^2J(\mathbf{R}_1)}{d^2\theta} = 4(k_{11}+k_{22})(3\sin^2\theta\cos^2\theta - \cos^4\theta) + 4(k_{12}+k_{21}) \times (3\sin^2\theta\cos^2\theta - \sin^4\theta). \tag{3}$$

[1] Note that here, we additionally assume that $\mathbf{A}$ is square and invertible.

Assume that the one-bit-matching condition is satisfied for $n = 2$ and $\nu_1^s > \nu_2^s$ and $k_1^m > k_2^m$; we have the two following possibilities for the matrix $\mathbf{K}$.

Case I. All $k_{ij} > 0$, that is, both sources are subgaussians or supergaussians. The necessary condition for local maxima is obtained by setting equation 2 = 0, and we have

$$\sin\theta\cos\theta = 0 \tag{4}$$

$$\tan^2\theta = \frac{k_{11} + k_{22}}{k_{12} + k_{21}}. \tag{5}$$

The roots for equation 5 correspond to local minima. Solving equation 4, we have $\hat{\theta} = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ which are local maxima as equation 3 < 0 at $\theta = \hat{\theta}$. It can be readily verified that each local maximum $\hat{\theta}$ leads $\mathbf{R}_1$ to a permutation matrix up to sign indeterminacy in this case.

Case II. $k_{11}, k_{22} > 0$ while $k_{12}, k_{21} < 0$, that is, a mixture that consists of one subgaussian and one supergaussian source. In this case, no real roots exist for equation 5. Among the roots $\hat{\theta} = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$, only 0 and π are local maxima as equation 3 < 0 there. The remaining two roots correspond to local minima. Similarly, it can be readily verified that each local maximum leads $\mathbf{R}_1$ to a permutation matrix up to sign indeterminacy in this case.

Summing the results of cases I and II, the proof is completed.

According to this theorem, when the skewness of each model probability density function is set to be zero, the two-source ICA problem can be successfully solved using any local descent algorithm of the typical objective function as long as the one-bit-matching condition is satisfied.

Acknowledgments

This work was supported by a grant from the Research Grant Council of the Hong Kong SAR (Project CUHK 4184/03E), and also by the Natural Science Foundation of China for Project 60071004.

References

Amari, S. I., Cichocki, A., & Yang, H. (1996). A new learning algorithm for blind separation of sources. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in Neural Information Processing*, *8* (pp. 757–763). Cambridge, MA: MIT Press.

Bell A., & Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*, 1129–1159.

Cardoso, J. F. (1999). Informax and maximum likelihood for source separation. *IEEE Signal Processing Letters, 4*, 112–114.

Liu, Z. Y., Chiu, K. C., & Xu, L. (2004). One-bit-matching conjecture for independent component analysis. *Neural Computation, 16*, 383–399.

Welling, M., & Weber, M. (2001). A constrained EM algorithm for independent component analysis. *Neural Computation*, *13*, 677–689.

Xu, L., Cheung, C. C., & Amari, S. I. (1998). Further results on nonlinearity and separation capability of a linear mixture ICA method and learned LPM. In C. Fyfe (Ed.), *Proceedings of the I&ANN'98* (pp. 39–45). Tenerife, Spain: NAISO Academic Press.

Received March 24, 2004; accepted June 22, 2004