Separability of Noisy ICA for High Dimensional Data

Yutaka Kano
Division of Mathematical Science, Graduate School of Engineering Science
Osaka University
Toyonaka, Osaka 560-8531, Japan
kano@sigmath.es.osaka-u.ac.jp

ABSTRACT: Independent component analysis (ICA, see, e.g., Hyvarinen, et al., 2001) is a technique of multivariate analysis that has been developed to separate a multivariate observational sensor vector **x** consisting of unobserved source signals **s** mixed linearly by an unknown mixed matrix *A*. The noisy ICA model is a variant of ICA models created by adding an error term **n**, represented as

$$\mathbf{x} = A\mathbf{s} + \mathbf{n},$$

where **x**, **s** and **n** are, respectively, *p*-, *m*- and *p*-dimensional random vectors with the zero mean vector, and **s** and **n** are independent. In addition, the components of **s** are independently and nonnormally distributed while **n** is assumed to be normally distributed. The (noisy) ICA model is said to be separable if the mixing matrix is identifiable and the original signals can be recovered from the observed sensors. In the ICA model i.e. $\mathbf{n} = 0$, Comon (1994) and Eriksson and Koivunen (2003) have proved that the ICA model is separable if *A* is of full column rank and at most one component of **s** may be normally distributed. On the other hand, conditions for identifiability and separability in the noisy ICA model have not been well developed. In this article, we first clarify several conditions for identifiability and separability (Davies, 2004), and point out inseparability of the noisy ICA model while the identifiability holds true under rather broad conditions. The inseparability of the noisy ICA model is serious because the main purpose of ICA is to recover the source signals. The phenomenon is similar to the factor indeterminacy problem in traditional factor analysis in psychology (Williams, 1978).
In this article, we suggest to augment sensor (observed) variables to achieve separability of the noisy ICA model. When whitening observed variables, we give mathematical conditions under which the model is going to be separable when sensor variables increase in number. A small numerical study is conducted to see how large *p* should be and how the parameter estimation influences upon the theory where the FastICA (Hyvarinen, 1999) is employed.

Keywords: *Factor Score Indeterminacy*, *PCA and FA*, and *Nonnormal and Independence*

References

COMON, P. (1994). Independent component analysis, A new concept? *Signal Processing*, **36**, 287-314.

DAVIES, M. (2004). Identifiability issues in noisy ICA. *Signal Processing Letters, IEEE*, **11**, 470-473.

ERIKSSON, J. and KOIVUNEN, V. (2003). Identifiability and separability of linear ICA models revisited. *Proceedings of the ICA2003*. 23-27.

HYVARINEN, A. (1999). Fast independent component analysis with noisy data using gaussian moments. In *Proc. Int. Symp. on Circuits and Systems*, Orlando, Florida,

HYVARINEN, A., KARHUNEN and OJA (2001). *Independent Component Analysis*. Wiley.

WILLIAMS, J. S. (1978). A definition for the common-factor analysis model and the elimination of problems of factor score indeterminacy. *Psychometrika*, **43**, 293-306.