Higher-order contrast functions improve performance of independent component analysis of functional MRI data with low signal-to-noise ratio

V. J. Schmithorst[1], and S. K. Holland[1]

[1]Pediatric Neuroimaging Research Consortium, Radiology, Children's Hospital Medical Center, Cincinnati, OH, United States

Introduction

Independent Component Analysis (ICA) is increasingly being used as a data-driven methodology for the analysis of functional MRI (fMRI) data [1]. The advantage of ICA is that it is not necessary to have prior knowledge of the hemodynamic response function (HRF). Thus, it is possible to use ICA to separate out cognitive networks without prior knowledge of their temporal dynamics. In addition, improved sensitivity is possible, compared to conventional hypothesis-driven analyses such as the General Linear Model [2], when the HRF is not specified with sufficient accuracy. ICA, however, requires a prior specification of the expected intensity distribution of the sources, denoted as a "contrast function" or "score function". In fMRI data sometimes sources may be present with low SNR. Here we present preliminary results suggesting that optimal performance in the presence of low SNR may involve the use of higher-order contrast functions.

Theory

Assuming an fMRI dataset *Y* contains *n* voxels and *T* time frames, the data is grouped into a *T*-X-*n* matrix. The noise-free ICA model is $Y = A\ S$ where *A* is a *T*-X-*T* mixing matrix and *S* is a *T*-X-*n* matrix of *n* spatially statistically independent sources ($S = W\ Y$ where *W* (the "unmixing matrix") = A^{-1}). In the standard Bayesian formalism [3], the objective function *O(W)* to optimize is $O(W) = \ln|\det(W)| + g(S)$, and thus the update rule on W is derived as $\Delta W = A^T + g'(S)Y^T$ where *g(S)* is the assumed log prior source density. Each voxel is treated as an independent observation randomly taken from the source density. (Typically *g(S)* is denoted as the "contrast function" while *g'(S)* is denoted as the "score function".) In the noiseless ICA case, the performance of the algorithm is robust to misspecification of the score function, provided the sources are correctly specified as super-Gaussian or sub-Gaussian [4]. A popular choice for the contrast and score function for superGaussian sources (present in fMRI data) is $g(S) = -\ln\cosh(aS), g'(S) = -a\tanh(aS)$, which closely resembles the contrast and score functions obtained from a Laplacian source distribution.

However, in the presence of noise this formalism is no longer valid, as it is impossible to obtain noise-free sources from pre-multiplication of the data matrix by the unmixing matrix. Estimation of the sources via optimization of joint likelihood is possible [5], but is computationally intensive, and is also dependent on accurate specification of the source densities. An alternative conceptualization of the source prior is to use, not an assumed source density, but a test statistic (such as kurtosis) and prior on that test statistic. In this case the only constraint on the score function for superGaussian sources is that the function monotonically increase with increasing superGaussianity (assuming constant variance). Kurtosis uses a higher power than 2 (it is the fourth moment), while the ln cosh closely resembles a linear function (order 1). Thus kurtosis optimizes for sources with outliers (long tails), while the ln cosh optimizes for sources with "spiky" peaks (many elements close to zero). This difference is significant as optimizing for long tails minimizes the chance of the algorithm becoming stuck in a local optimum and failing to optimally converge. It has been hypothesized that for fMRI data, higher-order contrast functions will outperform lower-order contrast functions [6], based on the specific shape of fMRI source distributions. Adaptive estimation of source densities has also been proposed [7]; however, this comes at the cost of a greater computational load.

Materials and Methods

Simulations were performed using routines written in IDL (Research Systems Inc., Boulder, CO). Datasets of 10,000 voxels and 25 time points were generated. Sources were generated from a Laplacian distribution. I.i.d. Gaussian noise was added, with a SNR of 0.7 and 1.5. ICA was performed, using both the ln cosh contrast function, and the kurtosis contrast function. A constrained optimization algorithm was used for ICA, utilizing gradient ascent; identical stopping criteria were used (mean(|dW|) < 1e-5). The sources found from the ICA were constrained to be orthogonal for comparison purposes (keeping det(W) constant), since kurtosis is not readily fit into a Bayesian framework. The ICA results were compared to the ground truth using the metric of the absolute value of the correlation coefficient. The correlation coefficients were averaged over the 25 sources.

Resting-state fMRI data was obtained from a normal adult volunteer on a Bruker 3T Medspec system. EPI-fMRI scan parameters were: TR/TE = 3000/52 ms, total scan time = 6 min., BW = 125 kHz, FOV = 25.6 X 25.6 cm, matrix = 64 X 64, slice thickness = 5 mm, 24 slices acquired covering the whole brain. Prior to the ICA decomposition, the data was pre-processed via spatial filtering using a Gaussian filter of width 4 mm, variance normalization of the time courses, and PCA reduction to 25 sources. The ICA analysis was performed as described above.

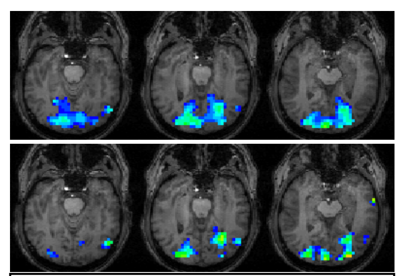

Figure 1. Comparison of ICA analysis of resting-state fMRI data using ln cosh contrast function (top) and kurtosis contrast function (bottom). Images thresholded at Z > 2.5, cluster size = 14 voxels.

Results and Discussion

Using kurtosis as the contrast function resulted in a significant improvement over the standard ln cosh function at lower SNR. The average correlation coefficient was 0.30 for kurtosis as compared to 0.25 for the ln cosh function when the SNR was 0.7 ($p < 0.05$, paired T-test); when the SNR was 1.5, the average correlation coefficient was 0.62 for kurtosis as compared to 0.60 for the ln cosh function ($p < 0.1$, paired T-test). For the resting-state fMRI data, for most sources the kurtosis and ln cosh functions showed high agreement ($R > 0.9$). However, for an occipital network there was not complete agreement ($R = 0.78$) between the two methods (Figure 1). Using kurtosis as the contrast function provided better localization of the activated regions within the occipital lobe, in agreement with previous results [6].

Further improvement is likely available using a metric which uses skewness of the source as well as kurtosis, since fMRI sources are often asymmetric as well as leptokurtic. We will investigate the use of a normality test, such as D'Agostino's K-squared test. Using such a test offers the advantage of re-formulating the ICA decomposition into a Bayesian framework by utilizing the confidence intervals on the test parameters.

Conclusion

The effect of using differing contrast functions in ICA of noisy fMRI data was compared. Results indicate using higher-order (>2) contrast functions such as kurtosis may provide superior performance than lower-order (<2) contrast functions.

References

[1] McKeown, M. J., Jung T.-P., Makeig S., et al. *Hum Brain Mapp*, 6, 1, 1998. [2] Worsley K. J., Friston K. J. *Neuroimage*, 2, 173, 1995. [3] MacKay D. J. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003. [4] Lee T. W., Girolami M., Sejnowski T. J. *Neural Comput*, 11, 41, 1999. [5] Hyvarinen A. *Neurocomputing*, 22, 49, 1998. [6] Suzuki K., Kiryu T., Nakada T. *Human Brain Mapping*, 15, 54, 2001. [7] Hong B., Pearlson G. D., Calhoun V. *Hum Brain Mapp*, 25, 297, 2005.