ICA of fMRI Group Study Data

Markus Svensén, Frithjof Kruggel

Max-Planck-Institute of
Cognitive Neuroscience
Leipzig, Germany
{svensen,kruggel}@cns.mpg.de

Habib Benali

INSERM U494
Paris, France
Habib.Benali@imed.jussieu.fr

Abstract

This paper presents a new approach for independent component analysis (ICA) of functional magnetic resonance imaging (fMRI) data, extending its applicability to simultaneous analysis of data from a group of subjects. This approach results in a set of timecourses common to all the subjects, combined with corresponding individual images for all subjects. This is in contrast to conventional treatment of group data, which normally involves per-voxel averaging of co-registered images over the group. Compared to conventional methods, the proposed methods maintains the effective spatial resolution of the fMRI data. The proposed method is illustrated using a real fMRI data set containing five subjects.

1 Introduction

Independent component analysis (ICA) is a signal processing technique for separating linear mixtures of source signals, which are assumed to be statistically independent. It has been subject to great interest within the neural information processing research community and analogies has been made between the sparse representations found by ICA and those formed in the brain.

McKeown et al. [8] proposed using ICA as a method for analysing functional brain imaging data of individual subjects, obtained using functional magnetic resonance imaging (fMRI). They suggested that ICA would separate out not only signals originating from the stimulation, which subjects receive during fMRI experiments, but also signals from other sources, such as slow varying sources and subject movements.

In this article, we consider extending this approach for simultaneously analysing fMRI data from a *group* of subjects. This yields a set of temporal patterns ('timecourses') common across the group, and for each timecourse, a separate image for each of the subjects. It is common practice to repeat a fMRI experiments across groups of subjects, but the subsequent analysis normally computes a per-voxel average over the group, which will reduce the effective spatial resolution and hide any individual differences.

In the following section we provide a very brief description of fMRI. This is followed by a section on ICA, its application to fMRI data, and how this can extended to deal with group data. In section 4, we give an example of our new approach applied to data from a real fMRI experiment including five subjects and the paper ends with a discussion.

2 Functional MRI

FMRI attempts to detect brain activity by localised, non-invasive measurements of the change in blood oxygenation, the so called *BOLD contrast* [9]. This is sensitive to the relative local concentrations of oxygenated hemoglobin (HbO_2) vs. deoxyhemoglobin and provides an indirect measure of the brain's neuronal activity. Measurements, in the form of a time-series of images, are collected under controlled conditions, where subjects are performing specific tasks, prompted by some stimulus (e.g. deciding whether a read out sentence is grammatically correct or not, perform arithmetic calculations, etc.).

The fMRI data are usually analysed by computing some sort of statistic, based on the correlation between the observed signals—the timeseries collected at each voxel—and a function representing the performance of the task after adjustment for hemodynamic effects (see e.g. [10]).

Most fMRI experiments are carried out with a group of subjects, all performing the same task sequence. In some cases, this is because the aim is to demonstrate a general effect on a population, rather than the effect on individuals. This is achieved by analysing the distribution of voxel-wise computed statistics over the group. In other cases, the signal from the stimulus induced activity is so low in comparison to signals from other sources, that the collected data needs to be averaged across subjects, in order for the signal to be detectable. In either case, the result is a statistical summary for the whole group, which does not describe individual differences and whose effective spatial resolution is significantly reduced compared to that of the the individual fMRI images, due to the high inter-

subject anatomical variance. At the moment, only few methods are available that simultaneously capture commonalities across a group and differences between individual subjects.

3 Independent component analysis

ICA [1, 2, 6] can be regarded as a generalisation of the more well-known principal components analysis [4]. Just like PCA, ICA linearly 'separates' a N-dimensional data set into N components. However, whereas PCA yields components which are uncorrelated, ICA tries to find components which are independent. These two conditions are identical when the components have Gaussian distributions, as is assumed by PCA [12], but ICA assumes that this is not the case. Given that the observed data was generated by linearly mixing independent sources with non-Gaussian distributions, ICA will separate the mixture into the original sources. It achieves this by computing a linear transform that 'unmixes' the mixed sources; if we think of the mixing transform as a matrix, ICA computes its inverse.

To make this more concrete, let $\boldsymbol{x}$ denote the N-dimensional random variable corresponding to the observed signals, and $\boldsymbol{s}$ the N-dimensional random variable corresponding to the underlying sources. We can then write

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}, \tag{1}$$

where $\boldsymbol{A}$ is a $N \times N$ full rank unknown *mixing matrix*. Note that we assume that the observed variables are observed without noise, but 'noise' may still be represented in the data through one or more of the sources.

Most ICA estimation procedures take their starting point in the assumed independence between the original sources, which simply means that the probability distribution function of the random vector $\boldsymbol{s}$ can be written as the product of marginal distributions of the vector elements

$$p(\boldsymbol{s}) = \prod_n^N p(s_n). \tag{2}$$

Based on this assumption the aim is to find an *unmixing matrix*, $\boldsymbol{W}$, such that the recovered sources

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} = \boldsymbol{W}\boldsymbol{A}\boldsymbol{s}, \tag{3}$$

are maximally independent. Several measures for independence have been proposed in the ICA literature and they can all be related to each other [2]. Here we restrict our attention to the mutual information, defined as

$$I(\boldsymbol{y}) = \sum_n^N H(y_n) - H(\boldsymbol{y}). \tag{4}$$

where

$$H(\boldsymbol{y}) = -\int p(\boldsymbol{y}) \ln p(\boldsymbol{y})\, d\boldsymbol{y}, \tag{5}$$

is the differential entropy for the random vector $\boldsymbol{y}$ with a probability distribution $p(\boldsymbol{y})$. We see directly from (2), (4) and (5) that the mutual information is zero when the recovered sources, $y_1, y_2, \ldots, y_M$, are independent and it can be shown to be strictly positive otherwise.

Assuming that $\boldsymbol{W}$ in (3) is of full rank, (4) can be written

$$I(y_1, y_2, \ldots, y_M) = \sum_n^N H(y_n) - H(\boldsymbol{x}) - \ln(\mathrm{abs}(\det(\boldsymbol{W}))). \tag{6}$$

If we consider minimising this expression with respect to $\boldsymbol{W}$, restricting its row vectors to be unit length, we note that the last term on the right hand side will be minimised when $\boldsymbol{W}$ is orthogonal, while the first term is minimised when the marginal distributions $p(y_m)$ are as 'non-Gaussian' as possible[1]. However, these marginal distributions are unknown—a complication which also affect all related measures of independence.

Several approximate schemes have been proposed, most of them primarily aimed at obtaining an estimate of the *gradient* of the objective function (e.g. (6) above) with respect to its parameters ($\boldsymbol{W}$ in (6)); these estimates are computed by averaging over a data set of samples from observed signals. Once the gradient has been obtained, the optimisation can be carried out using stochastic gradient or Newton methods.

3.1 ICA of fMRI Data

Applied to fMRI data, ICA tries to separate the sequence of recorded MR images into a set of independent source images. That means that each recorded image is treated as an observed variable, with the voxels in the image being regarded as sampled from that variable. McKeown et al. [8] motivate this with the argument that the spatial localisation of brain areas activated by performing the task in an fMRI experiment should be independent of the localisation of signals arising from sources such as head movement or system noise.

Thus, the mixing matrix ($\boldsymbol{A}$ in eq. (1)) will contain corresponding set of timecourses, specifying how the source images have been mixed to form the observed set of images. For example, we expect that the image(s) that represent the cognitive task has a timecourse that somehow reflects this. We obtain the mixing matrix directly as the inverse of the unmixing matrix found by ICA.

[1] For a random variable with given mean and (co)variances, the Gaussian distribution is the distribution that *maximises* the differential entropy.

3.2 ICA of fMRI Group Data

The idea of using ICA to analyse fMRI group data is based on the following observations:

- All subjects in an fMRI experiment are carrying out the same task sequence. Thus, the individual source images corresponding to the performance of the task ought to have similar timecourses.
- The union of two samples from N independent sources simply gives a larger sample where the N sources are still independent. This holds even when the two samples contains different sources.

These lead us to propose the following model (using the notation of eq. (1)):

$$[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_K] = \boldsymbol{A}[\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_K]. \qquad (7)$$

Here, $\boldsymbol{x}_k$ represent the data collected for subject k, in the form of a $N \times L_k$ matrix; N is the number of images collected for each subject during the experiment and L_k is the number of voxels inside the brain mask for subject k. Accordingly, $\boldsymbol{s}_k$ represent the matrix of independent source images of subject k. The $[\cdot, \cdot]$ operator denotes row-wise concatenation of matrices. $\boldsymbol{A}$, finally, denotes the mixing matrix, which is *common* to all subjects.

This means that we are extracting spatial components with common timecourses across all subjects. Compared to the approach of McKeown et al., we are also trying to obtain independent images, but the images now contain the voxels of all subjects. This approach does not require any coregistration of the images from different subject, since the spatial locations of the voxels are irrelevant to ICA. Thus, the resulting independent images has the same resolution as the original fMRI data.

Note that the ordering of the sources for an individual subject is arbitrary to ICA, and is determined only by the ordering of the columns in the common mixing matrix. Thus, we can readily assume that the task related source image(s) has the same 'index' (row index in the $\boldsymbol{s}_k$ matrices) across all subjects. Note also that, there is nothing preventing physical sources, such the task induced activation, to manifest themselves more than one source image, so this model can still cater for individual behaviour of such physical sources.

4 Example

We demonstrate our proposed method using fMRI data from an experiment with a ON/OFF block trial design and a visual alternating checkerboard stimulus [5]. The experiment was performed with five different subjects, from each of which data was collected from three slices. In addition to the functional data, a high resolution anatomical images was obtained for each slice of each individual.

We restrict our attention to a selected set of 228 functional images containing the first 9 blocks. We arrange the selected images from each subject as described in equation (7) and the resulting data matrix is passed to an ICA algorithm. For the examples presented here we have used the extended infomax algorithm [7], but have also obtained similar results with the FastICA algorithm [3]. By computing the correlation between the timecourses in the resulting mixing matrix and a function representing the performance of the task after hemodynamic correction, task related components can be identified [8]. The resulting spectrum of correlation scores dropped approximately exponentially, with the five top scores being 0.91, 0.57, 0.27, 0.26 and 0.20.

Component A in figure 1[2] shows the independent image (rotated 90° anti-clockwise) with the highest correlation with the task overlaid on the anatomical images of the five subjects. The corresponding timecourse is shown, together with the function representing the task, in the upper plot in figure 2. The top image in figure 1 are in good agreement with the Z-maps obtained by conventional analysis of this data set [5]. The highlit regions, situated in the visual cortex (V1), correspond to the region of the visual field which is activated during the ON-phase.

Component B in figure 1 shows another independent image whose timecourse, shown in the bottom plot in figure 2, is weakly anti-correlated (-0.17) with the stimulus. The highlit regions in this image correspond to the peripheral parts of the visual field. This component might reflect an attentional mechanism forcing the focus to the central visual field during the ON phase [11].

5 Discussion

This paper has introduced a new approach for ICA of fMRI data, extending its applicability to groups of subjects.This allows for detection of common temporal patterns, with individual spatial distributions in each of the subjects. This is achieved by restricting ICA to find a single unmixing matrix common to all subjects. Compared to conventional treatment of group data, our approach avoids direct averaging across subjects and thus produces images whose effective spatial resolution equals that of the original fMRI data. ICA as such does not require any a-priori specified timecourse, but identification of task related components will be easier if such a timecourse is available.

Our approach also has potential for discovering other sources which may show common temporal

[2] A colour version of this figure is presented in the electronic version of the proceedings.

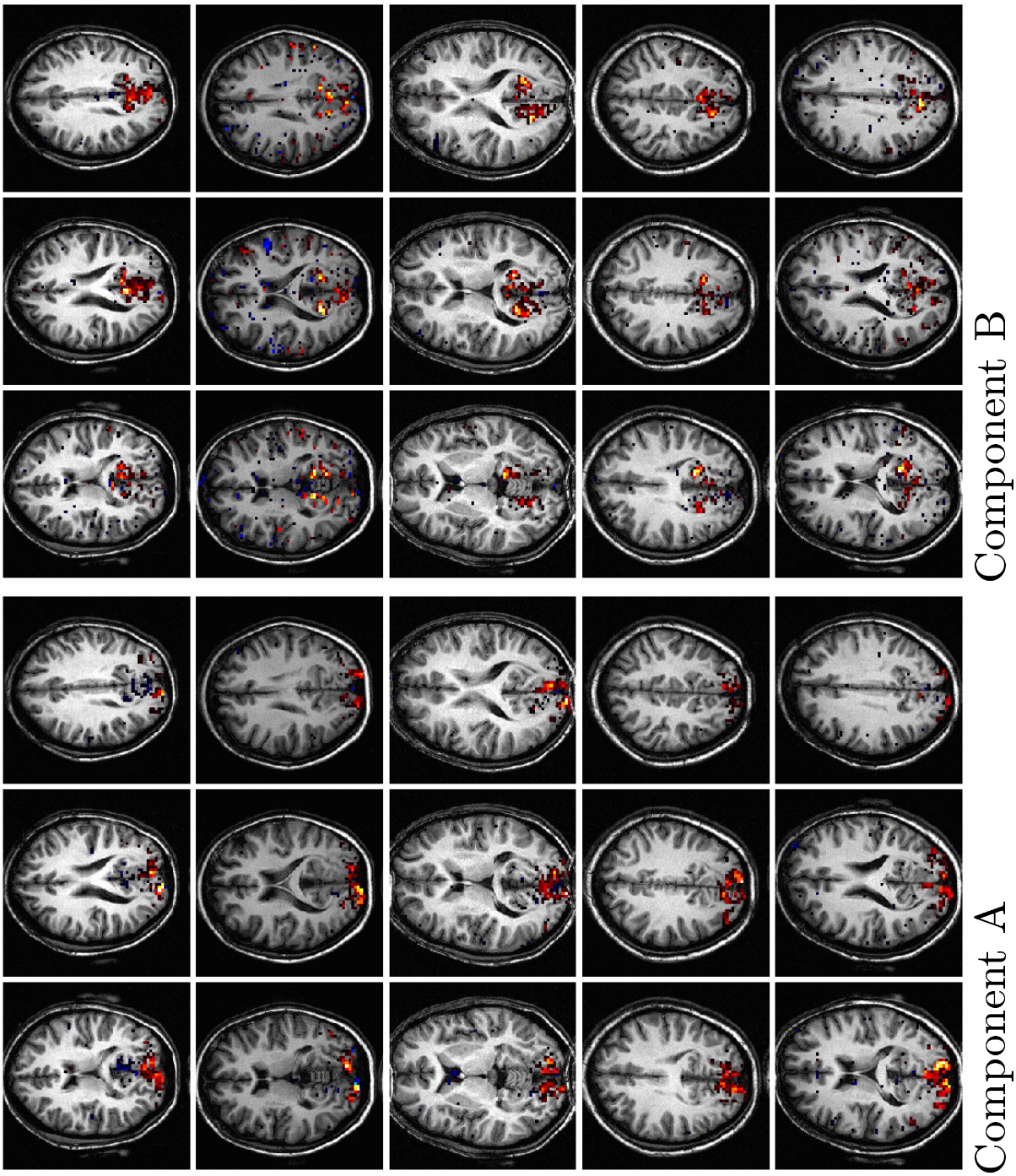

Figure 1: Two independent images (A and B) overlaid on the anatomical images of the five subjects. The images have been rotated 90° anti-clockwise. The corresponding timecourses are plotted in figure 2. The images have been normalised to zero mean and unit variance, and thresholded at ± 1.65. The blue-cyan voxels represent negative correlation, red-yellow voxels positive ditto and the brightness corresponds to the absolute value.

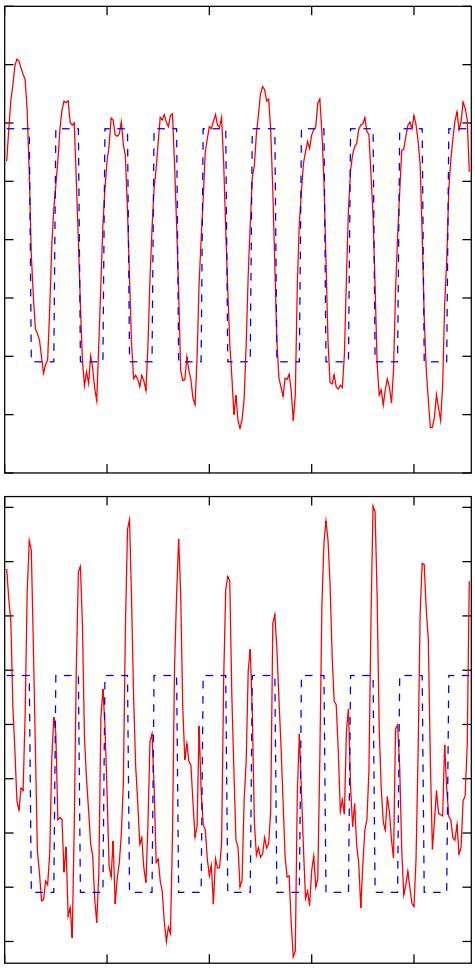

Figure 2: Timecourses (solid red) of the images shown in figure 1, plotted together with a time-shifted 'box-car' function (dashed blue) representing the performance of the task after hemodynamic correction. All timecourses have been normalised to zero mean and unit variance prior to plotting.

patters, such as attentional mechanisms, habituation or baseline drift in the fMR image acquisition system.

It lies in the nature of our approach that, it may suppress weak sources with different temporal characteristics across subjects. However, our model will still cater for strong sources represented in just one or a few subjects. For such sources, all voxels in images from subjects where the source is not present will have low values.

From a computational point of view, working with data from a group of subjects, will obviously require more computation than if we are working with data from a single subject from that group. However, the computational effort does not have to grow linearly. Our approach effectively means that we are increasing the number of samples from the observed variables, and hence the computation may converge with fewer iterations through the (larger) data set.

ICA of fMRI data is still a fairly new idea and many open questions still remain. As mentioned in section 4, several components with a stimulus related timecourse where found in our example analysis. This highlights the question on how to determine the contributions of individual components to the overall signal. Another problem, which is a difficult for the general case, is the identification of components that can aid our understanding of the generating processes of fMRI data, from the large set of components resulting from ICA. For the particular processes triggered by the stimulus, we can rely on correlation methods from conventional fMRI analysis. In the case of group data analysis, this clearly yields results in a new form, with a common timecourse extracted from the data and corresponding individual images for all subjects in the group.

Acknowledgement

The initial version of the code from which the results presented in this paper were obtained, was developed by A. Chojnowski.

References

[1] J.-F. Cardoso. Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 9(10):2009–2025, 1998.

[2] A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999. http://www.icsi.berkeley.edu/~jagota/NCS.

[3] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[4] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.

[5] F. Kruggel, C. J. Wiggins, C. S. Herrmann, and D. Y. von Cramon. Recording of event-related potentials during functional MRI at 3.0 tesla field strength. *Magnetic Resonance in Medicine*, 2000. In press.

[6] T.-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski. A unifying information-theoretic framework for independent component analysis. *Computers & Mathematics with Applications*, 1999. In press.

[7] T.-W. Lee, M. Girolami, and T. J. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, 11(2):417–441, 1999.

[8] M. J. McKeown, S. Makeig, G. G. Brown, T.-Z. Jung, S. S. Kindermann, T.-W. Lee, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI data by blind separation into indepedent spatial components. *Human Brain Mapping*, 6:160–188, 1998.

[9] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent blood oxygenation. *Proceedings of the National Academy of Sciences USA*, 87:9868–9872, 1990.

[10] J. C. Rajapakse, F. Kruggel, J. M. Maisog, and D. Y. von Cramon. Modeling hemodynamic response for analysis of functional MRI time-series. *Human Brain Mapping*, 6:283–300, 1998.

[11] A. S. Smith, K. D. Singh, and M. W. Greenlee. Attentional suppression of activity in the human visual cortex. *NeuroReport*, 11(2):271–276, 2000.

[12] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Statistical Methodology, Journals of the Royal Statistical Society, Series B*, 61(3):611–622, 1999.