

Probabilistic Independent Component Analysis in FMRI

C.F. Beckmann^{†‡} and S.M. Smith[†]

[†]Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Oxford, UK

[‡]Medical Vision Laboratory, Department of Engineering, University of Oxford, Oxford, UK

Abstract Independent Component Analysis (ICA) is classically performed using a square and noise-free mixing model, where the number of observations equals the number of source processes. In FMRI, both assumptions are known to be invalid and conflict with the model assumptions of the standard general linear model. We employ a probabilistic ICA (PICA) model for FMRI data that allows for more general non-square mixing in the presence of isotropic Gaussian noise and compare its performance against the conventional noise-free model.

In the probabilistic ICA model, the normalized (de-meant and variance-normalized) data time series $x(t)$ of length p are modelled as a mixture of q statistically independent unobserved source signals $s(t)$ which are linearly mixed by $A \in \mathbb{R}^{p \times q}$ and corrupted by additive noise $\nu(t)$ such that

$$x(t) = As(t) + \sigma\nu(t), \quad (1)$$

where $\nu(t) \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ denotes p -dimensional noise. Due to the presence of a noise term, the covariance matrix of the signal will be the sum of AA' and the noise covariance [2]

$$R_X = AA' + \sigma^2 I_p,$$

i.e. R_X will be of full rank and its rank will no longer equal the rank q of A . The signal subspace, however, will be spanned by the first q eigenvectors of R_X .

The choice of the number of components to extract is a problem of model order selection. Determining a cutoff value for the eigenvalues of an initial PCA using simple criteria like the reconstruction error or predictive likelihood [4] that do not incorporate a noise model will naturally predict that the accuracy steadily increases with increasing dimensionality. Instead, we employ Bayesian model order selection to infer the latent dimensionality using a Laplace approximation to the posterior distribution of the rank r within a probabilistic PCA model [5]. In order to account for the limited amount of data, we incorporate into this step the fact that due to the finite number of observations the sample covariance matrix $\hat{R}_\nu$ of the noise has a Wishart distribution. We utilise results from random matrix theory [2] on the empirical distribution function $G_n(\eta)$ for the eigenvalues of finite $p \times n$ dimensional random matrices: suppose that $p/n \rightarrow \gamma > 0$ as $n \rightarrow \infty$, then $G_n(\eta) \rightarrow G_\gamma(\eta)$ almost surely, where the limiting distribution has a density

$$g(\eta) = \frac{1}{2\pi\gamma\eta} \sqrt{(\eta - b_-)(b_+ - \eta)}, \quad b_- \leq \eta \leq b_+,$$

and where $b_\pm = (1 \pm \sqrt{\gamma})^2$. This allows to calculate the quantiles of the predicted cumulative distribution $G^{-1}(\eta)$, i.e. the expected eigenspectrum of a finite random Gaussian matrix. The latent dimension of the signal space is then estimated as in [5] and the data is projected into a signal + noise subspace spanned by the first k eigenvectors of the sample covariance matrix (whitened). Similarly, projecting the data onto the orthogonal eigenspace results in an estimate of the noise. Assumptions about the temporal smoothness of the source signals can be incorporated into this step by regularizing the eigenspectrum analysis, e.g by allowing for smooth principal components [6].

Within the lower-dimensional signal+noise subspace, independent components are identified by maximizing negentropy as a measure of non-Gaussianity [3]. Under the hypothesis of pure Gaussian noise, the regression coefficients in s are normally distributed and we can incorporate the uncertainty about the source estimates at each voxel location by using the voxel-wise standard deviation of the noise.

In order to determine the relevant voxels within the IC map, we fit a Gaussian mixture model (GMM) to the histogram of intensity values [1].

Example We acquired 180 volumes of FMRI data at 3T with an AB block design (30s on/off visual stimulation - b/w checkerboard flickering at 8 Hz). Fig. 1 shows the pattern of activation after applying GLM (FEAT www.fmrib.ox.ac.uk/fsl), PICA and noise-free ICA as in [4]. For both ICA and PICA, all maps are shown where the associated time course correlates with the GLM model at $r > 0.3$. In the case of ICA, the first IC depicts primary visual cortex but substantially differs from the GLM map. Additional signal appears fragmented into three different spatial maps. For PICA, the data was projected onto the first 23 eigenvectors prior to the ICA decomposition. The first IC shows an activation pattern much closer to the results from the GLM. A second IC highlights bilateral activation in what is probably V5/MT (visual motion perception).

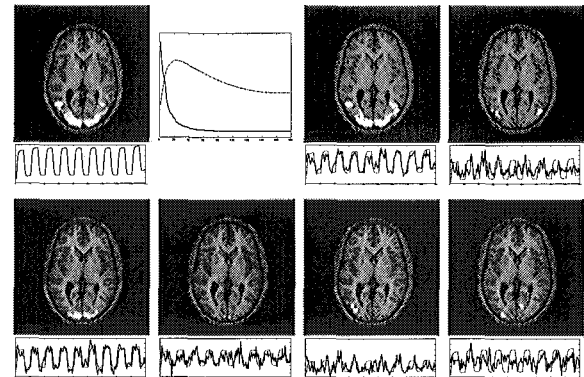


Fig 1: (top) Activation pattern from GLM analysis, eigenspectrum of the data together with Bayesian estimate of the dimensionality and IC maps from PICA; (bottom) IC maps from classical ICA analysis

Discussion We have outlined the basic principles of a probabilistic ICA model. Our results suggest that a probabilistic ICA model, where the number of sources is constrained, is less prone to overfitting, i.e. to the extraction of spurious spatial maps, while still allowing for the extraction of plausible activation patterns.

References

- [1] Beckmann, C.F. et. al. *NeuroImage* 13(6): S75&S76 (2001)
- [2] Everson, R., Roberts, S.J. *IEEE Tr. Sig. Proc.* 48(7): (1998)
- [3] Hyvärinen, A. *IEEE Tr. Neur. Netw.* 10(3): 626–634 (1999)
- [4] McKeown, M.J. *HBM* 6(3):160-188 (1998)
- [5] Minka, T. *NIPS* 13
- [6] Ramsay, J.O., Silverman B.W. *Functional Data Analysis*