Independent Component Analysis and Blind Source Separation

Aapo Hyvärinen

Dept of Mathematics and Statistics

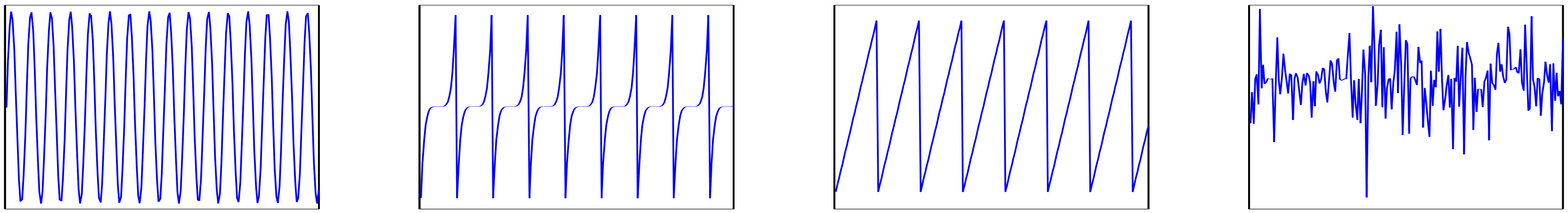
Dept of Computer Science and HIIT

University of Helsinki

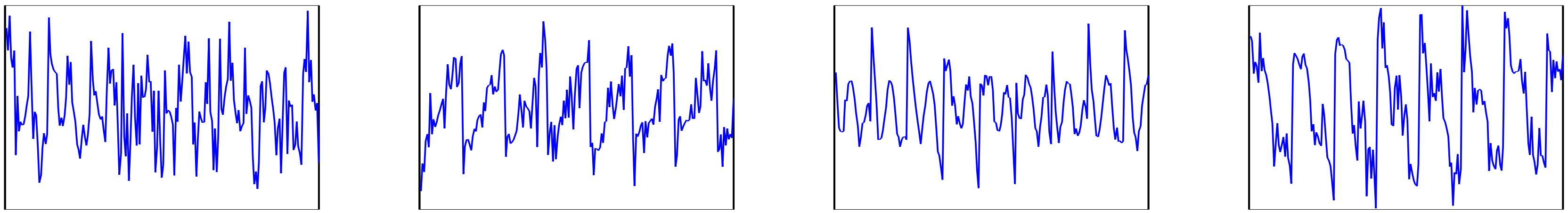
InterBrain Symposium 2010, Jyväskylä

Problem of blind source separation

There is a number of “source signals”:

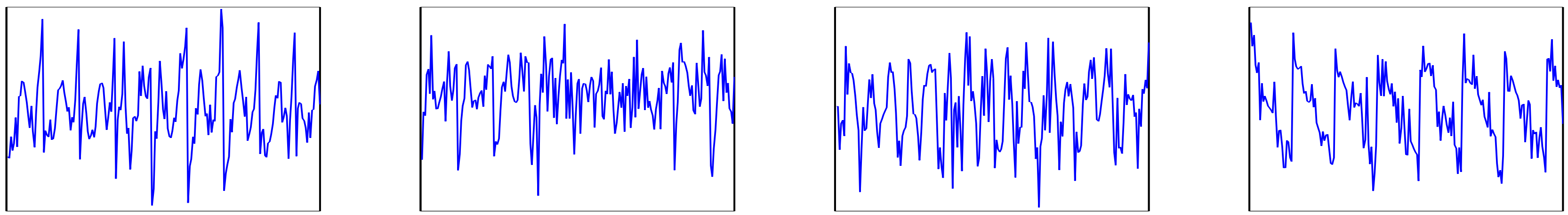


Due to some external circumstances, only linear mixtures of the source signals are observed.



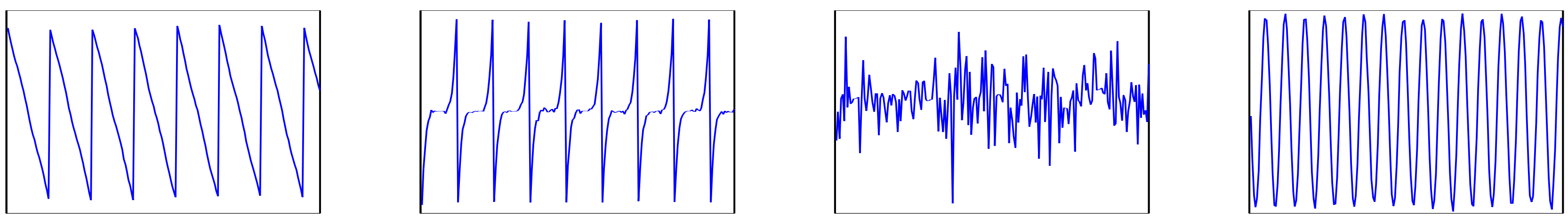
Estimate (separate) original signals!

Principal component analysis does not recover original signals



A solution is possible

Use information on statistical independence to recover:



Independent Component Analysis

(Hérault and Jutten, 1984-1991)

- Observed random variables x_i are modelled as linear sums of hidden variables:

$$x_i = \sum_{j=1}^{m} a_{ij} s_j, \qquad i = 1 \ldots n \tag{1}$$

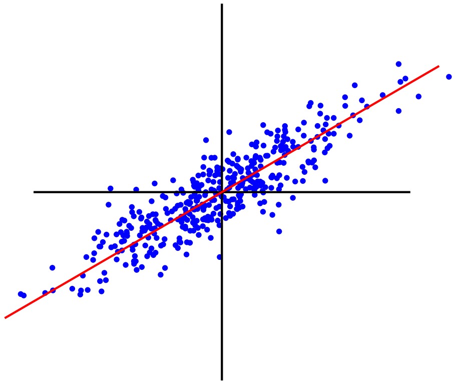
- Mathematical formulation of blind source separation problem
- Not unlike factor analysis
- Matrix of a_{ij} is constant (factor loadings), called “mixing matrix”.
- The s_i are hidden random factors called “independent components”, or “source signals”
- Problem: Estimate both a_{ij} and s_j, observing only x_i.

When can the ICA model be estimated?

- Must assume:
  - The s_i are mutually statistically independent
  - The s_i are nongaussian (non-normal)
  - (Optional:) Number of independent components is equal to number of observed variables
- Then: mixing matrix and components can be identified (Comon, 1994)
  A very surprising result!

Reminder: Principal component analysis

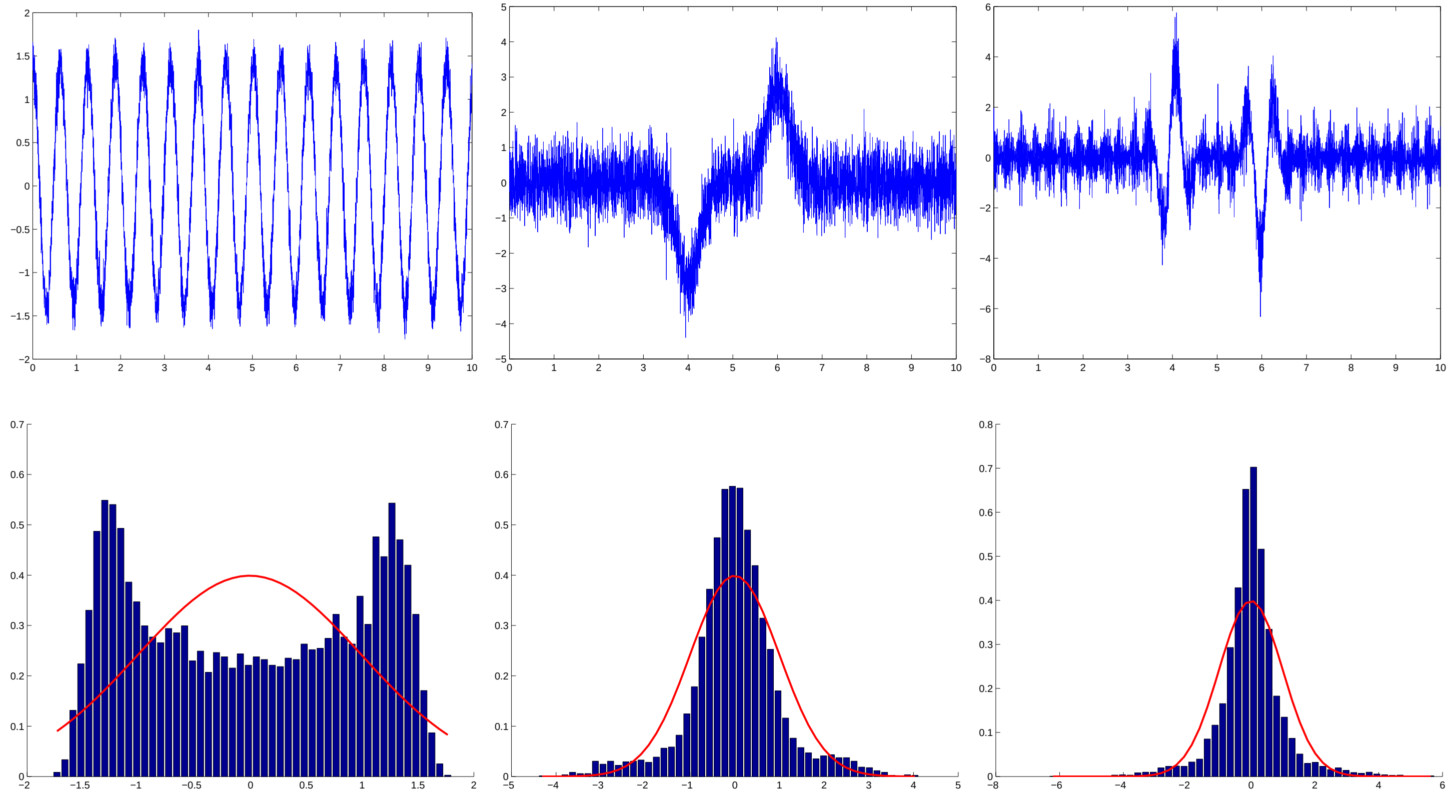
- Basic idea: find directions $\sum_i w_i x_i$ of maximum variance
- We must constrain the norm of $\mathbf{w}$: $\sum_i w_i^2 = 1$, otherwise solution is that w_i are infinite.
- For more than one component, find direction of max var orthogonal to components previously found.
- Classic factor analysis has essentially same idea as in PCA: explain maximal variance with limited number of components



Comparison of ICA, factor analysis and principal component analysis

- ICA is nongaussian FA with no separate noise or specific factors.
  So many components used that all variance is explained by them.
- No factor rotation left unknown because of identifiability result
- In contrast to FA and PCA, components really give the original source signals or underlying hidden variables
- Catch: only works when components are nongaussian
  - Many "psychological" hidden variables (e.g. "intelligence") may be (practically) gaussian because sum of many independent variables (central limit theorem).
  - But signals measured by sensors are usually quite nongaussian

Some examples of nongaussianity



Why classic methods cannot find original components or sources

- In PCA and FA: find components y_i which are uncorrelated

$$\operatorname{cov}(y_i, y_j) = E\{y_i y_j\} - E\{y_i\}E\{y_j\} = 0 \tag{2}$$

  and maximize explained variance (or variance of components)

- Such methods need only the covariances, $\operatorname{cov}(x_i, x_j)$
- However, there are many different component sets that are uncorrelated, because
  - The number of covariances is $\approx n^2/2$ due to symmetry
  - So, we cannot solve the n^2 factor loadings, not enough information! ("More equations than variables")
- This is why PCA and FA cannot find the underlying components (in general)

Nongaussianity, combined with independence, gives more information

- For independent variables we have

$$E\{h_1(y_1)h_2(y_2)\} - E\{h_1(y_1)\}E\{h_2(y_2)\} = 0. \quad (3)$$

- For nongaussian variables, nonlinear covariances give more information than just covariances.

- This is not true for multivariate gaussian distribution
  - Distribution is completely determined by covariances (and means)
  - Uncorrelated gaussian variables are independent, and their
  - distribution (standardized) is same in all directions (see below)

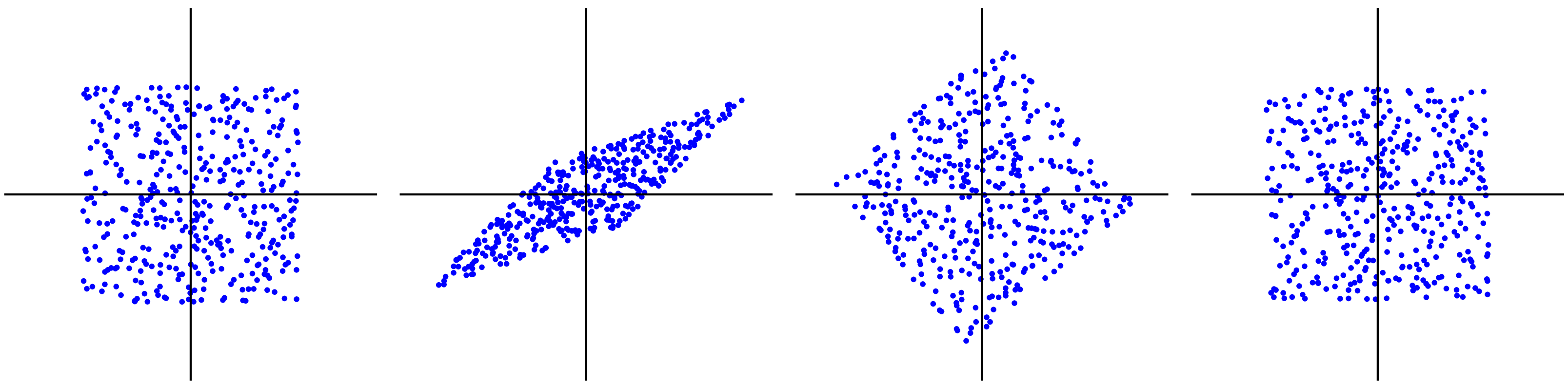
  $\Rightarrow$ ICA model cannot be estimated for gaussian data.

- In practice, simpler to look at properties of linear combinations $\sum_i w_i x_i$. PCA maximizes variance of $\sum_i w_i x_i$, can we do something better? Yes, see below.

Illustration

Two components with uniform distributions:

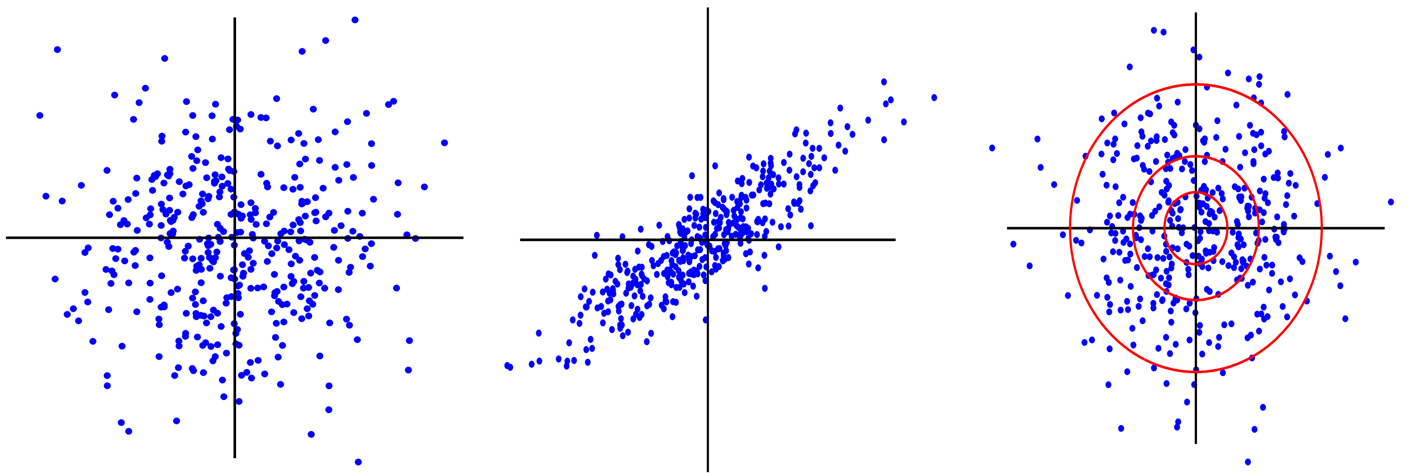
Original components, observed mixtures, PCA, ICA



PCA does not find original coordinates, ICA does!

Illustration of problem with gaussian distributions

Original components, observed mixtures, PCA



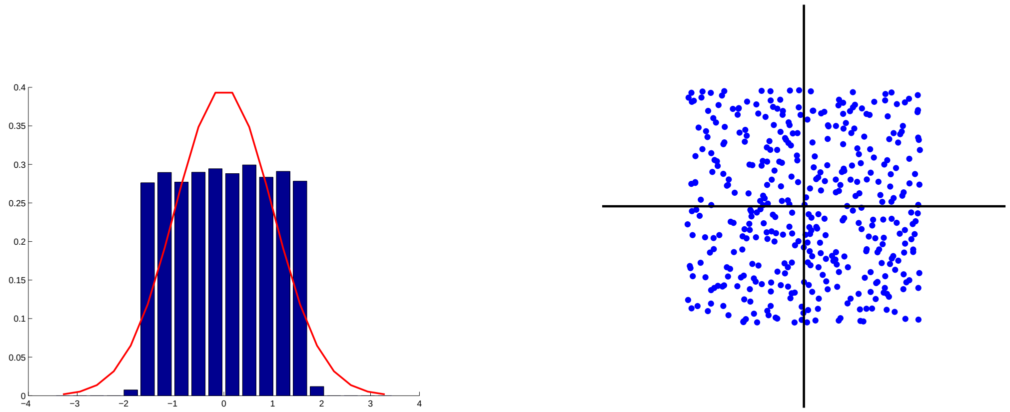
Distribution after PCA is the same as distribution before mixing!

“Factor rotation problem” in classic FA

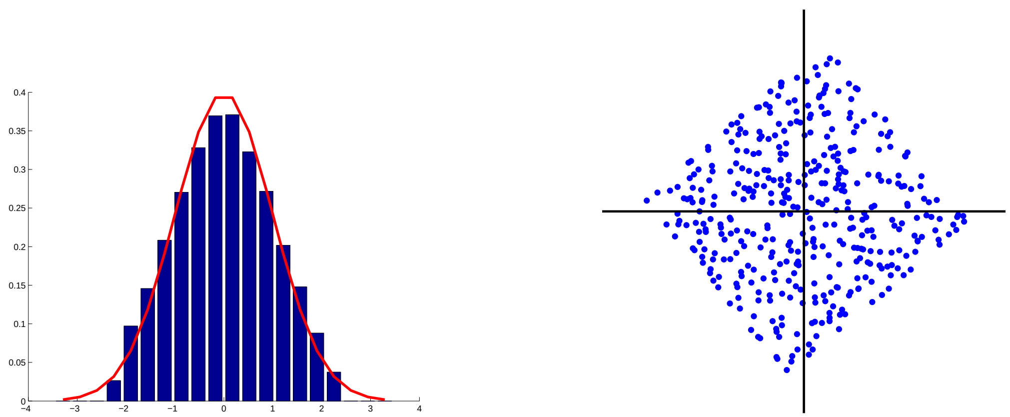
Basic intuitive principle of ICA estimation

- Inspired the Central Limit Theorem:
  - Average of many independent random variables will have a distribution that is close(r) to gaussian
  - In the limit of an infinite number of random variables, the distribution tends to gaussian
- Consider a linear combination $\sum_i w_i x_i = \sum_i q_i s_i$
- Because of theorem, $\sum_i q_i s_i$ should be more gaussian than s_i.
- *Maximizing the nongaussianity* of $\sum_i w_i x_i$, we can find s_i.
- Also known as projection pursuit.
- Cf. principal component analysis: maximize variance of $\sum_i w_i x_i$.

Illustration of changes in nongaussianity



Histogram and scatterplot, original uniform distributions



Histogram and scatterplot, mixtures given by PCA

Development of ICA algorithms

- Nongaussianity measure: Essential ingredient
  - Kurtosis: global consistency, but nonrobust.
  - Differential entropy: statistically justified, but difficult to compute.
    - Essentially same as likelihood (Pham et al, 1992/97) or infomax (Bell and Sejnowski, 1995)
  - Rough approximations of entropy: compromise
- Optimization methods
  - Gradient methods (e.g. natural gradient; Amari et al, 1996)
  - Fast fixed-point algorithm, FastICA (Hyvärinen, 1999)

Combining ICA with factor analysis or PCA

- In practice, it is useful to combine ICA with classic PCA or FA
  - First, find a small number of factors with PCA or FA
  - Then, perform ICA on those factors
- ICA is then a method of factor rotation
- Very different from varimax etc. which do not use statistical structure, and cannot find original components (in most cases)
- Reduces noise in signals, reduces computation

ICA on filtered data

- Temporal filtering possible: ICA still holds with the same matrix **A**

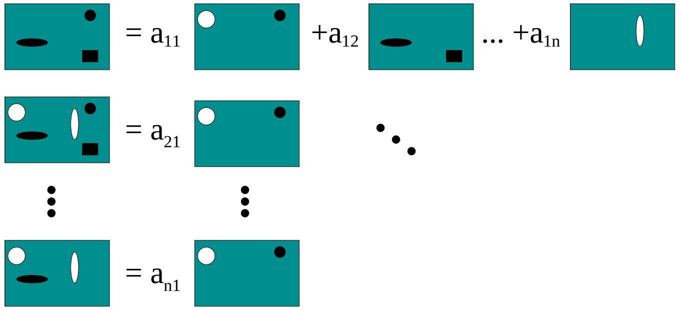
$$\tilde{x}_i(t) = f(t) * x_i(t) = \sum_\tau f(\tau) x_i(t-\tau) \tag{4}$$

$$\Rightarrow \tag{5}$$

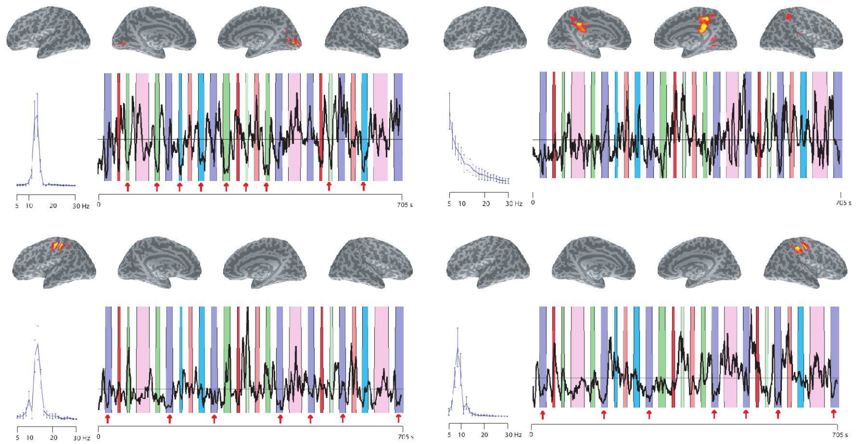
$$\tilde{x}_i(t) = \sum_j a_{ij} \tilde{s}_j(t) \tag{6}$$

- One can try to find a frequency band in which the source signals are as independent and nongaussian as possible
- Likewise, short-time fourier transform or wavelet transform can be done without changing **A**
- We argued (NeuroImage, 2010) that a short-time Fourier transform makes data more non-gaussian and improves source separation

Analysis of images: Spatial ICA

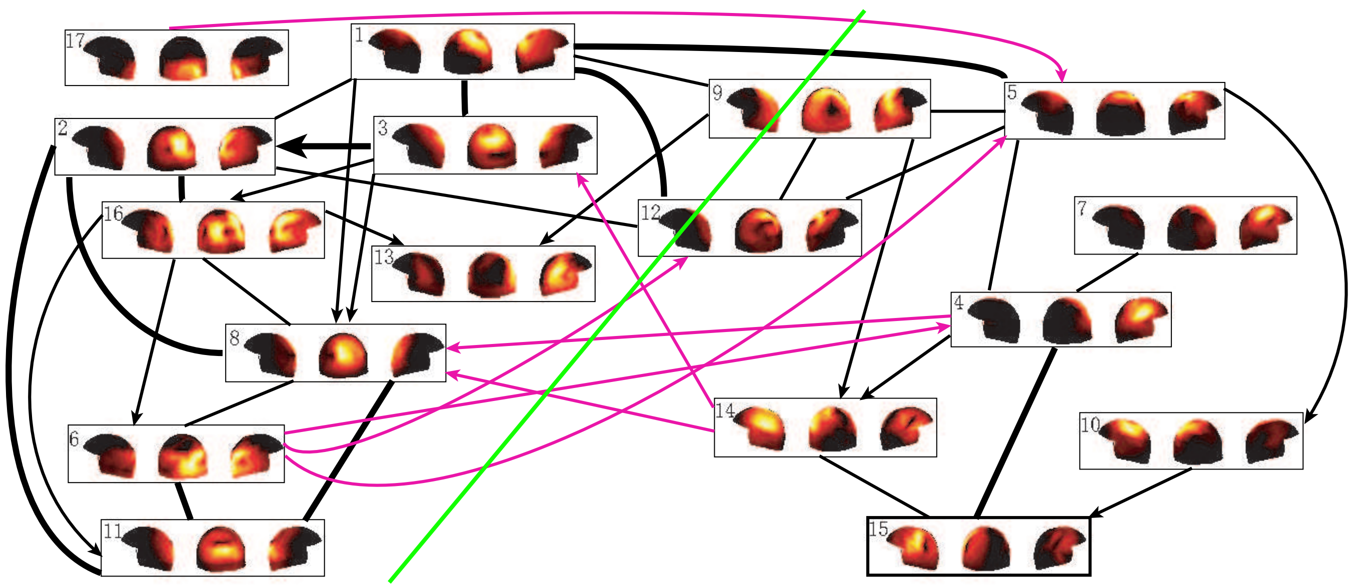


- Assume we observe several brain images at different time points
- ICA expresses observed images as linear sums of "source images"
- Reverses the roles of observations and variables
- Usually done only with fMRI, but after cortical projection also possible with MEG/EEG (Ramkumar et al, Biomag2010)



Connectivity (causality) analysis

- Gómez-Herrero et al (2008) combined ICA with a linear autoregressive model to analyze phase-coherence
- We formulate an autoregressive model based on energies/powers (Zhang & Hyvärinen, UAI 2010)



Black=positive, Red=negative, Green: manually inserted grouping

Reliability (significance) analysis

- Algorithmic reliability: Are there local minima?
- Statistical reliability: Is the result just a random effect?
- Can be analyzed by randomizing data, or initial point in optimization.
- Previous approaches randomize data from a single subject (Meinecke et al, 2002; Himberg et al, 2004)
- We proposed (HBM 2010) to consider inter-subject consistency
  - Do ICA separately for each subject
  - Formulate null hypothesis which says these are unrelated
  - Accept component only if it is found similar enough in a sufficient number of subjects

Conclusion

- ICA is very simple as a model:
  linear nongaussian latent variables model
- Solves factor rotation and blind source separation problems,
  if data (components) are nongaussian
- Estimate by maximizing nongaussianity of components
- Radically different from PCA both in theory and practice
- Recent and future work:
  - Short-time Fourier (or wavelet) transform to improve separation
  - Spatial ICA can be done if inverse operator available
  - Testing is an important but neglected topic
  - Connectivity analysis, next step?