Independent component analysis applied to EEG time series

Part 1: What is ICA?

ARNAUD DELORME
+ JOHN IVERSEN

33RD EEGLAB WORKSHOP LUBLIN

What is ICA?

Independent Component Analysis is a signal processing method to separate independent sources linearly mixed in several sensors.

ICA for dummies
http://arnauddelorme.com/ica_for_dummies

Bell and Sejnowski (1995) Infomax ICA
Makeig et al, (1995) First application of ICA to EEG
Motivation

Sensor-level EEG represents a mixture of cortical sources

We're interested in cortical activity
The ICA Linear Model

- Each source has a an associated “scalp map”, or pattern of electric potential measured at the electrodes

- As well as a “source activation” signal driving the map:

\[
s_i(t) \triangleq [a_1 a_2 \cdots a_n]
\]

- Notation: \(i\)th map is denoted, \(a_i\) and \(i\)th source activation, \(s_i\)

- EEG is linear superposition of independent sources:

\[
x(t) = a_1 s_1(t) + \cdots + a_n s_n(t) = As(t)
\]
$S = [a; b]$

Linear Combination

$X = AS$

**Independent component analysis**

Cocktail Party

Mixture of Brain source activity
Independent Component Analysis

ICA is a method to recover a version, of the original sources by multiplying the data by a unmixing matrix,

\[ S = WX, \]

- \( X \) is the data (channels x time)
- \( S \) are the ICA source activities (component x time)
- \( W \) is the ICA unmixing matrix (components x channels)

While PCA simply decorrelates the outputs (using an orthogonal matrix \( W \)), ICA attempts to make the outputs statistically independent, while placing no constraints on the matrix \( W \).
Plain English → MATLAB

Channel data = mixing (topo) * Source activation

EEG.data = EEG.icawinv * EEG.icaact

Source activation = unmixing * Channel data

Prior to ICA: sphere EEG channels data

EEG.icaact = (EEG.icaweights*EEG.icasphere) * EEG.data

[Note, EEG.icawinv = pinv(EEG.icaweights*EEG.icasphere) ]
Historical Remarks

ICA algorithms
- Bell & Sejnowski (1995): Information Maximization
- Amari et al. (1996): Natural Gradient Learning
- Cardoso (1996): JADE

Applications of ICA to biomedical signals
- EEG/ERP analysis (Makeig, Bell, Jung & Sejnowski, 1996).
- fMRI analysis (McKeown et al. 1998)

Part 2: How does Infomax ICA work?

Independent component analysis applied to EEG time series

33rd EEGLAB Workshop Lublin
ICA and PCA

ICA is a method to recover a version of the original sources by multiplying the data by a unmixing matrix.

While PCA simply decorrelates the outputs (using an orthogonal mixing matrix), ICA attempts to make the outputs statistically independent, while placing no constraints on the mixing matrix.

Components with overlapping scalp topographies

Central limit theorem

Scalp channels = linear mixture of A and B (more gaussian)

ICA Training Process

Sphering

Infomax
Super- and Sub-Gaussian sources

Kurtosis: a measure of how peaked or flat a probability distribution is; how non-gaussian.

\[ kurt(X) = \frac{E[(X - \mu)^4]}{\sigma^4} \]

Gaussian Dist. Kurtosis = 0
Super-Gaussian: kurtosis > 0 [biological]
Sub-Gaussian: kurtosis < 0 [oscillatory]

Infomax ICA maximizes kurtosis/non-gaussianity of sources

Entropy

\[ H(X) = -\sum_{x \in X} p(x) \log_b p(x). \]

\( p(x) = \) probability, \(-\log_2(p(x)) = \) information
Entropy = 'average information'
How informative is \( x \)?

Loaded Die (turns up 6 half of the time):
less uniform, less informative, lower entropy

\[ H = 5 \left( -\frac{1}{10} \log_2 \left( \frac{1}{10} \right) \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) = 2.16 \]
Entropy $H$

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_b p(x)$$

Joint entropy

$$H(X,Y) = - \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x,y) \log_b p(x,y)$$

Mutual Information $I$

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2).$$

How much does knowing $x$ tell you about $y$?
How different is joint distribution from project of marginals (departure from independence)

Shannon landmark 1948 paper "A Mathematical Theory of Communication."

ICA learning rule  
Bell & Sejnowski, 1995

How to make two variables statistically independent?

→ Minimize their mutual information.

Consider the joint entropy of two components:

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2).$$

Maximizing $H(y_1, y_2) \iff$ minimizing $I(y_1, y_2)$.

The learning rule:

$$\Delta W \propto \frac{\partial H(y)}{\partial W}$$

=0 if the two variables are independent

Entropy extremum
ICA in action

ICA in action

research.ics.tkk.fi/ica/icademo/
Independent component analysis applied to EEG time series

Part 3: ICA applied to EEG data

33rd EEGLAB Workshop Lublin

Makeig et al 1995

ICA/EEG Assumptions

- Mixing is linear at electrodes: OK
- Propagation delays are negligible: OK
- Components are non-gaussian: OK
- Component time courses are independent: ~
- Brain sources are stationary: ~
Key points of confusion

• ICA knows nothing about time
• ICA only maximizes *instantaneous* independence
• ICs can certainly have transient, or time-delayed correlations
  • Example: a choir singing in unison

Characteristics of Independent Components of EEG

- Artifacts
- Stimulus-locked activity
- Response-locked activity
- Non-phase locked activity
- Event-modulated oscillatory activity
- Overlapping Maps and Spectra
ICA applied to EEG
ICA applied to fMRI

Sample EEG ICA Decomposition

Eye blink
Fronto-parietal theta
Central theta
Occipital central alpha
Temporal muscle

ICA activity $U$

Temporal ICA
$X = W^{-1}U$
Inverse weight matrix $W^{-1}$
Data $X$ (EEG/MEG time series)

Spatial ICA
$X = W^{-1}U$
Inverse weight matrix $W^{-1}$
Data $X$ (fMRI voxel activities)
Part 4: A common use case: Removing ICA component artifacts in EEG

Independent component analysis applied to EEG time series

33rd EEGLAB Workshop Lublin
Artifact removal using ICA


Some Independent EMG Components
Independent component analysis applied to EEG time series

Part 5: We're really interested in the brain…
How do different ICA methods compare?

33rd EEGLAB Workshop Lublin

Validation of the ICA algorithm for EEG

Data
- 13 subjects performing a memory task
- 71 electrodes including EOGs
- more than 300,000 data points/subject

Decomposition
- 23 ICA algorithms plus PCA and Promax

Analysis
- Localization of all components with a single dipole (4-shell spherical model)

## Component examples

<table>
<thead>
<tr>
<th>Algorithm (Matlab func.)</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMICA</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>Pearson</td>
<td>ICentral (6)</td>
</tr>
<tr>
<td>Infomax (runica)</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>ERICA</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>SONS</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>SHIBBS</td>
<td>ICentral (5)</td>
</tr>
<tr>
<td>FastICA*</td>
<td>ICentral (2)</td>
</tr>
<tr>
<td>JADE (jader)</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>TICA</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>JADE optimized (jade_op)</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>JADE w/ time delay (jade_td)</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>eeA</td>
<td>ICentral (8)</td>
</tr>
<tr>
<td>Infomax (icaML) †</td>
<td>IC DTU Tbox</td>
</tr>
<tr>
<td>FObI</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>SOBIRO (ascobiro)</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>EVD 24</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>EVD</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>SOBI</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>icaMS†</td>
<td>IC DTU Tbox</td>
</tr>
<tr>
<td>AMUSE</td>
<td>ICALAB 1.5.2</td>
</tr>
<tr>
<td>PCA</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>Promax</td>
<td>EEGLAB 4.515</td>
</tr>
<tr>
<td>Whitening/Sphering</td>
<td>EEGLAB 4.515</td>
</tr>
</tbody>
</table>

Cumulative number of components for a given residual variance

Mutual information reduction (MIR)

For the linear transformation $y = Wx$, the entropy of the vector $y$ is given by,

$$h(y) = h(y_1, y_2, y_3, y_4, \ldots) = \log |\det W| + h(x)$$

The mutual information $I(y)$ is then,

$$I(y) = h(y_1) + \ldots + h(y_n) - h(y) = h(y_1) + \ldots + h(y_n) - \log |\det W| - h(x)$$

**Mutual information reduction (MIR)**, i.e. the amount of mutual information removed from set of channels, is given by,

$$\text{MIR} = I(x) - I(y) = [h(x_1) + \ldots + h(x_n)] - [h(y_1) + \ldots + h(y_n)] + \log |\det W| + h(x)$$

univariate
Infomax & AMICA

- Maximize independence of sources and
- Identify more dipolar (precisely localized) sources
  → More capable of identifying plausibly biological components

Correlations between decompositions

Available ICA implementations

- Infomax ICA
- runica matlab implementation
- binica compiled version; faster
- cudaica GPU version; fastest

- AMICA
- Delorme 2012: Best at extracting dipolar ICs
- Multiple-model support
Independent component analysis applied to EEG time series

Part 5: AMICA = Adaptive Mixture ICA

33rd EEGLAB Workshop Lublin

Features of AMICA

- Can model arbitrary source distributions (as mixtures)
- Can accommodate non-stationarity by decomposing data into multiple ICA models

- AMICA plugin (https://sccn.ucsd.edu/~jason/amica_help.html)
- Post-AMICA utilities plugin (https://github.com/sccn/postAmicaUtility)
Modeling Unknown Source Densities – Generalized Gaussian Densities

- The Generalized Gaussian density has the following basic form, where \( \rho \) is the “shape parameter”:

\[
\mathcal{G}(s; \rho) \triangleq \frac{1}{2 \Gamma(1 + 1/\rho)} \exp(-|s|^\rho)
\]

- Adding location and scale parameters:

\[
\mathcal{G}(s; \mu, \beta, \rho) \triangleq \frac{\beta^{1/2}}{2 \Gamma(1 + 1/\rho)} \exp(-\beta^{\rho/2}|s - \mu|^\rho)
\]

Generalized Gaussian Mixtures

- More complex densities can be constructed using a mixture model:

\[
p_M(s) = \sum_{j=1}^{m} \alpha_j p_j(s)
\]

- For example, consider, the following model:

\[
\frac{1}{2} \mathcal{G}(s; -2, 1, 1) + \frac{2}{15} \mathcal{G}(s; 0, 1, 2) + \frac{3}{10} \mathcal{G}(s; 2, 1, 10)
\]
Segmentation over Tasks

- 3 models learned on 1.5 hour recording with multiple tasks
  - Continuous Performance Task (CPT), Eriksen Flanker, Fast Response, Eyes Closed (EC), Eyes Open (EO)

Hsu et al., Modeling brain dynamic state changes with adaptive mixture independent component analysis, NeuroImage 2018

- Sleep stage classification
- 8 Models, unsupervised, segmenting REM, Deep Sleep, and Wake states
Computationally costly, so use NSG!

- People have often not used AMICA because it is time-consuming to compute and requires a HPC cluster, ideally
- High power computing is now freely available to everyone, using the Neuroscience Gateway
- nsgportal plugin for EEGLAB (Ramón)
- https://www.nsgportal.org

Independent component analysis applied to EEG time series

Part 6: Reproducibility of ICA decompositions

33rd EEGLAB Workshop Lublin
Do components reproduce in single subjects?

- 2 subjects
- 11 datasets per subjects (acquired on different days)
- Data preprocessed
- Informax ICA run on each dataset
- Match components using Corrmaph plugin of EEGLAB


Cluster 11

Relica: A method for estimating the reliability of ICA components

Artoni, et al., Neuroimage, 2014
**Within-cluster reliability**

The distribution of dipolarity within the cluster helps assessing the quality and characteristics of Independent Components.

**Reliability criteria and the rv<15%**

First justification why we should select an rv <15% for components to include in further analyses: there is a forbidden region underlined in red, that indicates the absence of

---

Independent component analysis applied to EEG time series

**Part 7: Common misconceptions about ICA and open questions**

33rd EEGLAB Workshop Lublin

---

**Does ICA always work?**

**No**
ICA is only useful to remove artifacts in EEG... No

ICA adulterates the phase of the EEG signal... No

https://youtu.be/BfqCh2UeJik

https://sccn.ucsd.edu/wiki/ICA_phase_distortion

About 50 messages on EEGLABLIST
https://eeglablist/2017/012642.html
ICA components are independent, so how can you do connectivity analysis?...no problem!

ICA should NOT be used for connectivity analysis of underlying brain sources...No

Remember: ICA only finds instantaneous independence

See my lecture tomorrow on Connectivity Analysis

ICA is the only way to identify brain sources...No

Other methods for distributed source exist: LORETA, beamforming
e.g. Haufe et al 2010.
Interpreting ICA components as brain sources is compelling, but may not ideally capture non-stationary, moving sources (traveling waves) ... Yes

• Alpha activity can exhibit dependence and coherence
• Alpha as a 'distributed dynamic' phenomenon

Residual mutual information following ICA decomposition – dependent subspaces

S. Makeig, R. Martinez-Cancillo, 2018

B = brain
M = muscle
E = eye
? = other
SC = channel
Outstanding questions

*Need your help to answer these!*

- How account for inter-individual variations in components?
- Develop new methods to characterize non-stationary dynamics, dependent subspaces
- Develop new methods to regularly assess reliability of ICA decompositions

---

**Articles about ICA and EEG**


Delorme, A., Palmer, J., Oostenveld, R., Onton, J., and Makeig S. (2012) Linear EEG decompositions that are more independent include more near-dipolar components. *PLOS One*, 7(2).
