Source Separation for Biomedical Signals: Blind or not Blind?

Christian Jutten

GIPSA-lab, CNRS UMR 5216, Univ. of Grenoble and Inst. Univ. of France

Contents of the lecture

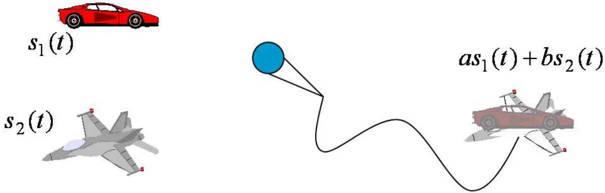
I. The source separation principles

Part 1 Part 2

Source separation : a usual problem?

Example

The signal at sensor (electrode, antenna, microphone, etc.) is a mixture of signals



Question

Is it possible to restore the various signals (called sources) only from the mixture observed at the sensor output?
If yes, under what conditions? And how?

Source separation with priors

Observation : $s_1(t) + s_2(t)$

Classical approaches

- With frequency priors: signal and noise non overlapping frequency bands $\Rightarrow$ simple filtering.
- Widrow and Hoff approach (1960) based on a (noise) reference:
  - One has $G[s_2]$, where G is an unknown filter;
  - one estimates a filter $F[.]$ so that $F[G[s_2]]$ cancel the noise contribution in the mixture
  - F such that : $E[(s_1(t) + s_2(t) - F[G[s_2]])s_2(t)] = 0$

Without priors...

What can we do ?

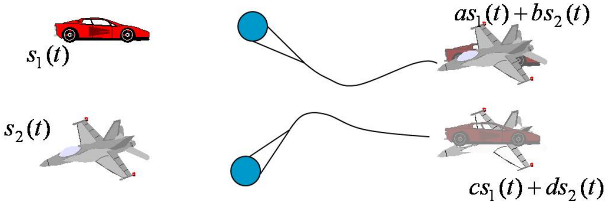
- if signal and noise are in the same frequency range?
- if one has no noise reference?

Source separation: the fondamental idea

A few observations

Separation becomes possible if one has a few observations:

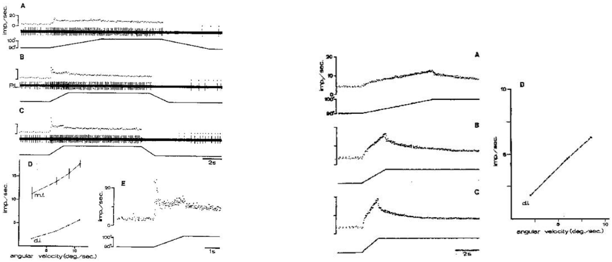
- more sensors than sources,
- different (non proportional) mixtures, i.e. $ad \neq bc$



It is the spatial **diversity**

Origin of the problem

Motion decoding in vertebrate [HJA85]



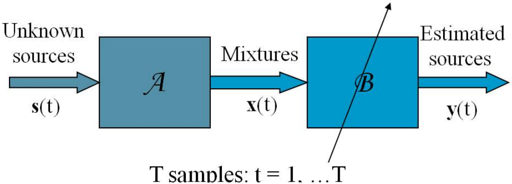
$$f_I(t) = a_{11}p(t) + a_{12}v(t)$$
$$f_{II}(t) = a_{21}p(t) + a_{22}v(t).$$

$p(t)$ = joint position ; $v(t)$ = joint speed

Questions ?

- Can one recover $p(t)$ and $v(t)$ from the mixtures?

Mathematical formalization



Assumption on unknown mixtures

Unknown mixtures are invertible (more sensors than sources):

- linear instantaneous or convolutive mixtures,
- nonlinear mixtures.

Principles of the solution

- Direct: Estimate $\mathcal{A}$ only from the mixtures (observations)
- Indirect: Separating block, $\mathcal{B}$, suited to mixing block $\mathcal{A}$.
- Are $\mathcal{A}$ or $\mathcal{B}$ identifiable? How estimate them?

Ill-posed problem without extra priors on sources

No solution if...

Darmois's result (1953) [Dar53]

Linear Factorial Analysis

- Linear mixture: $\boldsymbol{x} = \mathbf{A}\boldsymbol{s}$
- Assumption: components of the random vector $\boldsymbol{s}$ are mutually independent

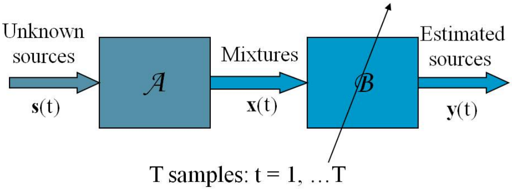
Theoretical result

- Separation is impossible if sources are **independent and identically distributed (iid) AND Gaussian**

Two directions for separation

- If sources are **iid AND NON Gaussian**. *ICA, with HOS*
- If sources are **NON iid and Gaussian** *with SOS*:
  - temporally correlated sources (colored signals)
  - non stationnary sources

Independent Component Analysis (ICA)



Principle

- General case: $\boldsymbol{x} = \mathcal{A}(\boldsymbol{s})$, with $P = K$ and $\mathcal{A}$ invertible
- Assumption: sources are statistically independent
- Idea: estimate $\mathcal{B}$ such that $\boldsymbol{y} = \mathcal{B}(\boldsymbol{x})$ has statistically independent components

Questions

- If $\boldsymbol{y}$ is independent $\Leftrightarrow \mathcal{B} = \mathcal{A}^{-1}$?
- Estimation of $\mathcal{B}$ based on independence of $\boldsymbol{y}$. Criterion?

ICA: linear instantaneous mixtures

Mixing model

- $\boldsymbol{x} = \mathbf{A}\boldsymbol{s}$

A theoretical result (Comon, HOS 1991 and SP 1994) [Com94]

- Let $\boldsymbol{x}(t) = \mathbf{A}\boldsymbol{s}(t)$, where $\mathbf{A}$ is a regular matrix and $\boldsymbol{s}(t)$ is a source vector with statistically independent components, with at most one is Gaussian, then $\boldsymbol{y}(t) = \mathbf{B}\boldsymbol{x}(t)$ is a random vector with mutually independent component if and only if $\mathbf{BA} = \mathbf{DP}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{P}$ is a permutation matrix.

Comments

- Independence $\Leftrightarrow$ separation, with weak assumptions: $\mathbf{A}$ regular matrix, non Gaussian and independent sources
- Global mapping $\mathbf{BA} = \mathbf{DP} \Rightarrow$ Scale and permut. ambiguities
- At most one Gaussian source

ICA: convolutive linear mixtures

Mixing model

- $\boldsymbol{x}(t) = \mathbf{A}(t) * \boldsymbol{s}(t)$ or in discrete time representation:
- $x_i(n) = \sum_j (\sum_k a_{ij}(k) s_j(n-k)), \forall i = 1, \dots, K$

First theoretical results [YW94, TJ95]

- Let $\boldsymbol{x}(t) = [\mathbf{A}(z)]\boldsymbol{s}(t)$, where $\mathbf{A}(z)$ is an invertible matrix whose entries are filters, and $\boldsymbol{s}(t)$ is a source vector with statistically independent components, with at most one is Gaussian, then $\boldsymbol{y}(t) = [\mathbf{B}(z)]\boldsymbol{x}(t)$ is a random vector with mutually independent component if and only if $\mathbf{B}(z)\mathbf{A}(z) = \mathbf{D}(z)\mathbf{P}$, where $\mathbf{D}(z)$ is a diagonal filter matrix and $\mathbf{P}$ is a permutation matrix.

Comments

- Independence $\Leftrightarrow$ separation, up to a filter and a permutation

ICA: nonlinear (NL) mixtures

Mixing model

- $\mathbf{x}(t) = \mathcal{A}(\mathbf{s}(t))$

Generally impossible [Dar53, HP99]!

- In the NL case, independence of $\mathbf{y}$ does not imply source separation
- $\mathbf{y}(t) = \mathcal{B}(\mathbf{x}(t))$ can have independent components with a non diagonal (i.e. mixing) global mapping $\mathcal{B} \circ \mathcal{A}$!

Possible for particular mixing mappings

- Post-nonlinear mixtures [TJ99, BZJN02, JBZH04, AJ05]
- NL mixtures which can be linearized [KLR73, EK02]
- Bilinear and linear-quadratic mixtures (without identifiability proof: [HD03, DH09])

Blind, actually blind?

Source separation is a problem...

- non blind: mixing nature is known, assumption on sources
- in ICA, *blind* $\leftrightarrow$ *unsupervised*, in reference to *blind* equalization or *blind* deconvolution

Examples of priors on sources

- Sources with temporal dependence between samples: NON iid
- Discrete-valued or bounded sources: algebraic or geometric methods
- Sparsity: Sparce Component Analysis (SCA)
- Positiviy: nonnegative factorization

Source separation in other representation domains?

For linear mixtures

- Consider the linear mixture: $\boldsymbol{x}(t) = \mathbf{A}\boldsymbol{s}(t)$
- Transform with any mapping $\mathcal{T}$, which preserves linearity leads to another linear mixtures:

$$\tilde{\boldsymbol{x}}(t) = \mathcal{T}(\boldsymbol{x}(t)) = \mathcal{T}(\mathbf{A}\boldsymbol{s}(t)) = \mathbf{A}\mathcal{T}(\boldsymbol{s}(t)) = \mathbf{A}\tilde{\boldsymbol{s}}(t)$$

- Examples of such mappings: wavelet transform, Fourier transform, DCT, etc.
- Solve the source separation problem in the new space: $\hat{\tilde{\boldsymbol{s}}}(t)$
- Come back to the initial space, with inverse of $\mathcal{T}$: $\hat{\boldsymbol{s}}(t) = \mathcal{T}^{-1}(\hat{\tilde{\boldsymbol{s}}}(t))$

Determined or overdetermined linear mixtures

Determined mixtures

- Equal numbers of sources and sensors (mixtures), $K = P$
- $\mathbf{A}$ regular matrix

Overdetermined mixtures

- More sensors (mixtures) than sources, $K > P$
- Solution if the mixing matrix is full rank (P)
- Pre-processing with PCA for projecting data in the P-dimension signal subspace before separation

In both cases, one can use...

- identify the mixing matrix $\mathbf{A}$ or the separation matrix $\mathbf{B}$, provides the sources

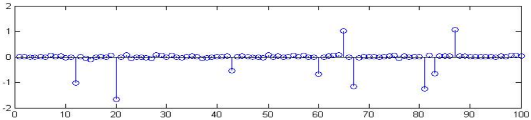
Underdetermined mixtures

More sources than sensors (mixtures), $P > K$

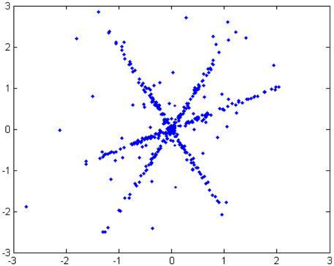
- If $\mathbf{A}$ is known (its inverse does not exist!), one cannot directly estimate $\boldsymbol{s}$
- Identification of $\mathbf{A}$ and source estimation are two distinct and tricky problems
- Without extra priors, infinite number of solutions [TJ99]
- Possible solution for discrete or sparse sources

Why are sparse sources interesting?

- Sources may be sparse in time or in any domain, preserving the linear nature of mixtures



- 2 mixtures of 3 sparse sources



Sparse Component Analysis

- A framework for solving underdetermined source separation
- Review paper (see [GL06] and Chapter 10 in [CJ10]),
- Initial underdetermined source separation problem:
  $\boldsymbol{x}(t) = \mathbf{A}\boldsymbol{s}(t)$, nbr sources $P \gg K$ nbr d'observations

Main ideas

1. Transform with a sparsifying mapping $\mathcal{T}$, which preserves linearity (e.g. wavelet transform, ST Fourier transform, etc.):

$$\tilde{\boldsymbol{x}}(t) = \mathcal{T}(\boldsymbol{x}(t)) = \mathcal{T}(\mathbf{A}\boldsymbol{s}(t)) = \mathbf{A}\mathcal{T}(\boldsymbol{s}(t)) = \mathbf{A}\tilde{\boldsymbol{s}}(t)$$

2. Solve the source separation problem in the sparse space: $\hat{\tilde{\boldsymbol{s}}}(t)$
3. Come back to the initial space, with inverse of $\mathcal{T}$:
   $\hat{\boldsymbol{s}}(t) = \mathcal{T}^{-1}(\hat{\tilde{\boldsymbol{s}}}(t))$

Separation in noisy mixtures

For linear mixtures

- Noisy mixtures: $\boldsymbol{x}(t) = \mathbf{A}\boldsymbol{s}(t) + \boldsymbol{n}(t)$, where the noise $\boldsymbol{n}$ is independent of $\boldsymbol{s}$.
- Noise has two main effects:
  1. it leads to errors in estimating $\mathbf{B}$
  2. even if $\mathbf{B}$ is perfectly estimated, i.e. $\mathbf{B} = \mathbf{A}^{-1}$, then:

$$\boldsymbol{y}(t) = \mathbf{B}\boldsymbol{x}(t) = \boldsymbol{s}(t) + \mathbf{B}\boldsymbol{n}(t)$$

- Perfect source separation $\not\Rightarrow$ best output SNR

ICA methods

For iid NON Gaussian sources, ICA algorithms are based on independence criteria

Random vectors

- Let $\mathbf{S}$ be a random vector with P components $S_1, \ldots S_P$
- Components are mutually statistically independent iff $p_{S_1 S_2 \ldots S_P}(u_1, u_2, \ldots u_P) = p_{S_1}(u_1) p_{S_2}(u_2) \ldots p_{S_P}(u_P)$

Comments

- Criterion tricky to use
- Relation between 2 multivariate functions

Kullback-Leibler divergence

KL divergence as independence criterion

- $KL(p_{\mathbf{Y}} \| \Pi p_{Y_i}) = \int \ldots \int p_{\mathbf{Y}}(\boldsymbol{u}) \log \frac{p_{\mathbf{Y}}(\boldsymbol{u})}{\Pi p_{Y_i}(u_i)} d\boldsymbol{u}$
- $KL(p_{\mathbf{Y}} \| \Pi p_{Y_i})$ is a positive real number, which vanishes iff components of the randon vector $\mathbf{Y}$ are independent
- This measure, $KL(p_{\mathbf{Y}} \| \Pi p_{Y_i})$, is egal to mutual information (MI) $I(\mathbf{Y})$ well known in information theory
- Drawback: computation of KL divergence requires marginal and joint pdf's!
- Advantage: good independence measure, always positive, which vanishes iff independence
- Practically, approximations of densities lead to different constrast functions.

Decomposition of MI with entropies

Definition

$$I(\mathbf{Y}) = \int \ldots \int p_{\mathbf{Y}}(\boldsymbol{u}) \log \frac{p_{\mathbf{Y}}(\boldsymbol{u})}{\prod p_{Y_i}(u_i)} d\boldsymbol{u} \tag{1}$$

$$= \sum_i H(Y_i) - H(\mathbf{Y}) \tag{2}$$

with the following definitions for marginal et joint entropies

$$H(Y_i) = \int p_{Y_i}(u_i) \log p_{Y_i}(u_i) du_i \tag{3}$$

$$H(\mathbf{Y}) = \int \ldots \int p_{\mathbf{Y}}(\boldsymbol{u}) \log p_{\mathbf{Y}}(\boldsymbol{u}) d\boldsymbol{u} \tag{4}$$

Problems

- Estimation of joint and marginal entropies requires joint and marginal densities

A trick for linear mixtures

In the linear determined case, with an invertible mixing matrix, $\mathbf{A}$, since $\mathbf{Y} = \mathbf{B}\mathbf{X}$, one can write:

The trick

$$I(\mathbf{Y}) = \sum_i H(Y_i) - H(\mathbf{Y}) \tag{5}$$

$$= \sum_i H(Y_i) - H(\mathbf{X}) - E[\log \mid \det \mathbf{B} \mid] \tag{6}$$

Consequence

- $\min_{\mathbf{B}} I(\mathbf{Y}) \Leftrightarrow \min_{\mathbf{B}} \sum_i H(Y_i) - \log \mid \det \mathbf{B} \mid$
- The trick avoids estimating joint entropy

MI Minimisation and score function

For solving linear mixtures, one estimates a separating matrix, $\mathbf{B}$, which minimizes $I(\boldsymbol{y})$.

Derivative of MI with respect to B

$$\frac{dI(\mathbf{Y})}{d\mathbf{B}} = \sum_i \frac{dH(Y_i)}{d\mathbf{B}} - \frac{d \log \mid \det \mathbf{B} \mid}{d\mathbf{B}}$$

$$= \sum_i E[\underbrace{\frac{-d \log p_{Y_i}(y_i)}{dy_i}}_{\varphi_{y_i}(y_i):\ \text{score function of } y_i} \frac{dy_i}{d\mathbf{B}}] - \mathbf{B}^{-T}$$

MI Minimization and HOS

After some algebra: $\frac{dI(\mathbf{Y})}{d\mathbf{B}} = 0 \Leftrightarrow E[\varphi_{\mathbf{Y}}(\mathbf{Y})\mathbf{Y}^T] = \mathbf{I}$

minimizing $I(\mathbf{Y}) \Leftrightarrow$ zeroing high-order moments $E[\varphi_{y_i}(y_i)y_j] = 0$

Independence or decorrelation?

Decorrelation ?

- Independence $\Rightarrow$ decorrelation, but reverse is wrong
- However, decorrelation is a first step to independence, hence 2-step approaches with whitening or sphering

Algebraic point of view

- For P linear mixtures of P sources (assumed zero mean, iid)
  - $\mathbf{B}$ has P^2 unknown parameters,
  - using SOS, one has $P + P(P-1)/2 = P(P+1)/2$ equations related to $E(Y_1^2), \ldots E(Y_P^2)$ and $E(Y_1 Y_2), E(Y_1 Y_3) \ldots, E(Y_P Y_{P-1})$

A 2-step approach

For real linear mixtures

- **B** can be factorized in 2 matrices:

  - a whitening (or sphering) matrix **W**
  - an orthogonal matrix **U**
- **W** is computed so that $E[\mathbf{Z}\mathbf{Z}^T] = \mathbf{I}$, i.e. $E[\mathbf{WAS}(\mathbf{WAS})^T] = \mathbf{WA}E[\mathbf{SS}^T](\mathbf{WA})^T = \mathbf{WA}(\mathbf{WA})^T = \mathbf{I}$
- It means that **WA** is an orthogonal matrix, consequently **U** must be an orthogonal matrix too.
- **W** is a symmetric matrix, and its estimation requires $P(P+1)/2$ parameters
- **U** is associated to $P(P-1)/2$ plane (Givens) rotations

Decorrelation not enough for iid Gaussian

After whitening

- samples $z(t)$ are spatially non correlated: $E[\mathbf{ZZ}^T] = \mathbf{I}$
- for Gaussian iid sources, after whitening, sources are statistically independent: there is no way for estimating $\mathbf{U}$!

Using mutual information

- Minimizing Mutual Information: $E[\varphi_{y_i}(y_i)y_j] = 0, \forall i \neq j$, where φ_{y_i} is the score function
- For Gaussian (iid or not) sources with unit variance, $\varphi_{y_i} = +y_i$
- Consequently, for Gaussian iid variables, $E[\varphi_{y_i}(y_i)y_j] = E[y_i y_j]$: MI minimisation is equivalent to decorrelation!

Second order separation

For iid sources

- Impossible!

For non iid sources

- Colored sources: AMUSE, Tong et al.[TSLH90]; SOBI, Belouchrani et al. [BAM93], Molgedey and Schuster [MS94]
- Nonstationary sources: Matsuoka et al. [MOK95]; Pham and Cardoso [PC01]

Basic idea

Idea

- Computing a separating matrix $\mathbf{B}$ which decorrelates simultaneously $y_i(t)$ and $y_j(t-\tau)$, for various τ, and $\forall i, j$
- Equivalent to cancel simultaneously $E[y_i(t)y_j(t-\tau)]$, $\forall i \neq j$, i.e. diagonalize simultaneously $E[\boldsymbol{y}(t)\boldsymbol{y}^T(t-\tau)]$, for at least two values of τ.

Joint diagonalization

- Covariance matrices of $\boldsymbol{s}$, $\mathbf{R}_{\boldsymbol{s}}(\tau) = E[\boldsymbol{s}(t)\boldsymbol{s}^T(t-\tau)]$: diagonal
- Covariance matrices of $\boldsymbol{x}$:
  $\mathbf{R}_{\boldsymbol{x}}(\tau) = E[\boldsymbol{x}(t)\boldsymbol{x}^T(t-\tau)] = \mathbf{A}\mathbf{R}_{\boldsymbol{s}}(\tau)\mathbf{A}^T$, $\forall\tau$
  can be simultaneously diagonalized by a matrix $\mathbf{B}$

Identifiability theorem for colored sources

Theorem

- The mixing matrix $\mathbf{A}$ is identifiable from second order statistics (up to scale and permutation indeterminacies) iff the correlation sequences of all sources are pairwise linearly independent, i.e. if $(\rho_i(1), \ldots, \rho_i(K)) \neq (\rho_j(1), \ldots, \rho_j(K))$, $\forall i \neq j$

In frequency domain

- Due to uniqueness of Fourier transform, the identifiability condition in the time domain can be transposed in the frequency domain:
  The mixing matrix $\mathbf{A}$ is identifiable from second order statistics (up to scale and permutation indeterminacies) iff the sources have distinct spectra i.e. pairwise linearly independent spectra.

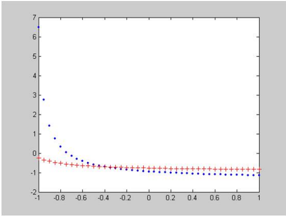
Consequences

- Since the separation is achieved using second order statistics (SOS), Gaussian sources can be separated
- Since we just use SOS, maximum likelihood approaches can be developped assuming Gaussian densities.

Non stationary sources

Covariance matrices on 2 time windows

- Assume $\boldsymbol{x}(t) = \mathbf{A}\boldsymbol{s}(t)$, with $\mathbf{A}$ regular and with nonstationary sources, with pairwise variance ratio $(\sigma_i/\sigma_j)^2(t) \neq \mathrm{cst}$
- On 2 time windows with different variance ratios: one get 2 different covariance matrices, $\mathbf{R}_1(0)$ and $\mathbf{R}_2(0)$,
- Diagonalizing $\mathbf{R}_i(0)$ alone leads to infinite nbr of solutions
- Jointly diagonalizing $\mathbf{R}_1(0)$ and $\mathbf{R}_2(0)$ leads to unique solution
- 2 mixtures, 2 sources: $(\sigma_2/\sigma_1)^2 = 1$, $(\sigma_2/\sigma_1)^2 = 2$.



Non stationary sources (con't)

Pham and Cardoso approach [PC01]

- Time is divided into p time blocks, W_i with length T_i,
- Covariance matrix on each window is estimated using

$$\hat{\mathrm{R}}_{\boldsymbol{x},i} = \frac{1}{T_i} \sum_{t \in W_i} \boldsymbol{x}(t)\boldsymbol{x}^T(t), \ \forall i = 1 \ldots, p$$

- Assuming Gaussian sources (SOS method), Pham and Cardoso [PC01] prove that ML estimate of the separating matrix $\mathbf{B}$, for unknown source variances, is provided by joint diagonalization of the estimated matrices $\hat{\mathrm{R}}_{\boldsymbol{x},i}$, provided that source variance ratios are not constant.

Joint diagonalization: exact or approximate ?

Exact joint diagonalization (EJD)

- EJD of 2 symetric or positive-definite matrices exists
- In practice, due to matrix estimation errors:
  - EJD does not exist for more than 2 matrices
  - EJD of 2 matrices sensitive to matrix estimation errors

Approximate joint diagonalization (AJD)

- AJD of a set of matrices more robust than EJD of 2 matrices
- How to choose the set of matrices (delays $\tau_1, \ldots \tau_K$, windows, etc.) for the best estimation ?
- AJD is a usual problem in data analysis $\Rightarrow$ huge number of fast and efficient algorithms

II. BIOMEDICAL APPLICATIONS

Part 1 Part 2

How to design a source separation method?

- Many methods for solving source separation problems
- Estimation methods, which require:
  - a separating model, suitable to the mixing model,
  - a criterion: source independence (ICA), source decorrelation (for non iid), etc.
  - an optimisation algorithm.
- When running source separation algorithm, one obtains a solution, which is the best one under the above constraints.

Are the Extracted Sources relevant ?

For chosing a relevant algorithm

Modelling the observations

How the physical system provides the observations?

- It leads to a mixing model, and hence to a relevant model of separating structure.
- Modelling is important for showing what should be the estimated sources.
- Physics of the system is important for component interpretation.

Choose the best representation domain for applying BSS

Use all the possible priors

- Positivity, sparsity, temporal dependance, periodicity or cyclostationnarity, discrete-valued data, etc.
- Priors can lead to better criteria than independence.

Source separation of biomedical signals

Two main questions

For biomedical signals...

- What is the nature of the mixtures ?
  - linear instantaneous: with electric/magnetic sources (EEG, ECG)
  - convolutive: when sources are acoustic sources (e.g heart murmur)
  - nonlinear: very few exemples
- What are the sources properties?
  - iid, or with time structures (colored, non stationary),
  - in a given frequency range (alpha, etc.)
  - particular shape (ECG)

Source separation on MEG

MEG recordings

- With different tasks done by the subject

R. Vigário, E. Oja / Neural Networks 13 (2000) 891–907 897

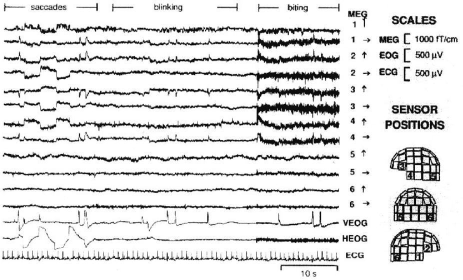


Fig. 6. A subset of 12 spontaneous MEG signals from the frontal, temporal and occipital areas. The data contain several types of artifacts, including ocular and muscle activity, the cardiac cycle, and environmental magnetic disturbances. Adapted from Vigário, Jousmäki et al. (1998).

Source separation on MEG

IC extracted with FastICA

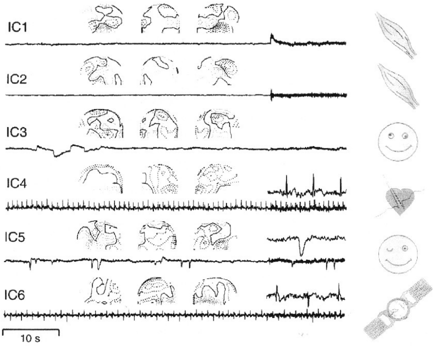


Fig. 7. Artifacts found from MEG data, using the FastICA algorithm. Three views of the field patterns generated by each independent component are plotted on top of the respective signal. Full lines correspond to magnetic flux exiting the head, whereas the dashed lines correspond to the flux inwards. Adapted from Vigário, Jousmäki et al. (1998).

Joint diagonalisation for EEG

Using priors

- Sources are colored: variance-covariance matrices with various delays
- Sources are nonstationary: variance-covariance matrices on different windows
- Frequency priors: sources in particular bands, e.g. Alpha, Beta, Theta, etc.

In frequency or wavelet domain

- The linear model still holds, after short term Fourier or wavelet transforms
- AJD can be done on cospectral matrices in the Fourier domain

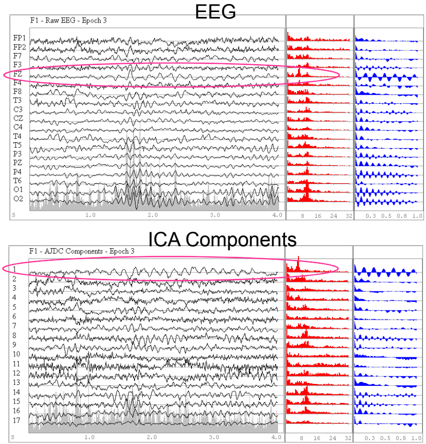
AJD possible with various matrices

- since the mixing matrix $\mathbf{A}$ is preserved

Source separation with frequency priors on EEG

Theta rythm sources extracted by AJD

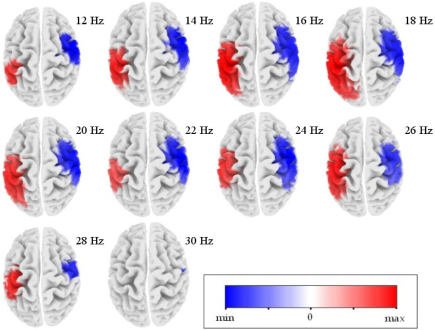
With linear filtering, $\boldsymbol{x}(t) = \mathbf{A}\boldsymbol{s}(t)$ becomes $\tilde{\boldsymbol{x}}(t) = \mathbf{A}\tilde{\boldsymbol{s}}(t)$



Source separation for BCI

Localisation according to the frequency range

- On each frequency band j, $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ becomes $\tilde{\mathbf{x}}_j(t) = \mathbf{A}\tilde{\mathbf{s}}_j(t)$
- IC computed by AJD on various frequency bands, for L/R hand motion



General Eigen Value Decomposition (GEVD) approach

GEVD problem

- Observation: $\boldsymbol{x}(t) = \mathbf{A}\boldsymbol{s}(t) + \boldsymbol{n}(t) = \boldsymbol{x}_s(t) + \boldsymbol{x}_n(t)$, where $\boldsymbol{x}_s$ and $\boldsymbol{x}_s$ are desired and undesired signals, resp.
- Objective: estimate the spatial filter $\boldsymbol{b}$ such that $y = \boldsymbol{b}^T\boldsymbol{x}$ has desired property
- Maximizing the power ratio according to 2 conditions can be associated to the joint diagonalization of two matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, called GEVD problem (Rayleigh-Ritz theorem):

$$\max_{\mathbf{B}} \frac{\mathbf{B}^T\mathbf{M}_1\mathbf{B}}{\mathbf{B}^T\mathbf{M}_2\mathbf{B}} \Leftrightarrow \mathbf{B} \text{ such that } \mathbf{M}_1\mathbf{B} = \mathbf{M}_2\mathbf{B}\mathbf{D}$$

where $\mathbf{D}$ is the eigenvalue diagonal matrix

General Eigen Value Decomposition (GEVD) problem (con't)

SNR maximizer

$$\max_{\boldsymbol{b}} SNR(\boldsymbol{b}) = \max_{\boldsymbol{b}} \frac{E[y_s^2]}{E[y_n^2]} = \max_{\boldsymbol{b}} \frac{\boldsymbol{b}^T \mathbf{R}_{\boldsymbol{x}_s} \boldsymbol{b}}{\boldsymbol{b}^T \mathbf{R}_{\boldsymbol{x}_n} \boldsymbol{b}} \Leftrightarrow \boldsymbol{b} = GEVD(\mathbf{R}_{\boldsymbol{x}_s}, \mathbf{R}_{\boldsymbol{x}_n})$$

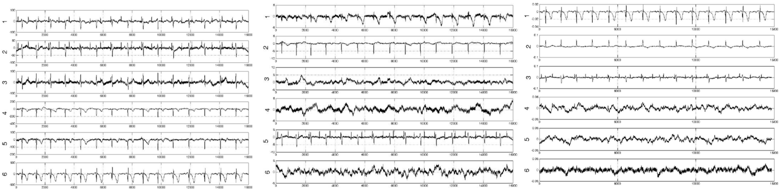
General Eigen Value Decomposition (GEVD) problem (con't)

Periodicity maximizer: Sameni et al. [TSS+09, SJS10]

- $$\min_{\boldsymbol{b}} \frac{E_t[(y(t+\tau)-y(t))^2]}{E[y(t)^2]} \Leftrightarrow \boldsymbol{b} = GEVD(\mathbf{R}_{\boldsymbol{x}}(\tau), \mathbf{R}_{\boldsymbol{x}}(0))$$

- Application for ECG extraction: component are sorted by similarity to ECG (R peaks)

Mixtures | Est. sources ICA (JADE) | Est. sources PiCA



General Eigen Value Decomposition (GEVD) problem (con't)

Spectral-contrast maximizer

- $$\max_{\boldsymbol{b}} \frac{E_{f \in BP}[|\mathbf{Y}(f)|^2]}{E_f[|\mathbf{Y}(f)|^2]} = \max_{\boldsymbol{b}} \frac{\boldsymbol{b}^T \mathbf{S}_{\boldsymbol{x},BP} \boldsymbol{b}}{\boldsymbol{b}^T \mathbf{S}_{\boldsymbol{x}} \boldsymbol{b}} \Leftrightarrow \boldsymbol{b} = GEVD(\mathbf{S}_{\boldsymbol{x},BP}, \mathbf{S}_{\boldsymbol{x}})$$
- Application for extraction of signal with prior that power in important in a given frequency band, BP. This idea has been used for atrial fibrillation extraction [PZL10, LISC11]

General Eigen Value Decomposition (GEVD) problem (con't)

Nonstationary maximizer

$$\max \frac{E_{t\in W_1}[\boldsymbol{y}^2(t)]}{E_{t\in W_2}[\boldsymbol{y}^2(t)]} = \max_{\boldsymbol{b}} \frac{\boldsymbol{b}^T \mathbf{R}_{\boldsymbol{X},W_1}(0)\boldsymbol{b}}{\boldsymbol{b}^T \mathbf{R}_{\boldsymbol{X},W_2}(0)\boldsymbol{b}} \Leftrightarrow \boldsymbol{b} = \mathrm{GEVD}(\mathbf{R}_{\boldsymbol{X},W_1}(0), \mathbf{R}_{\boldsymbol{X},W_2}(0))$$

Common spatial pattern

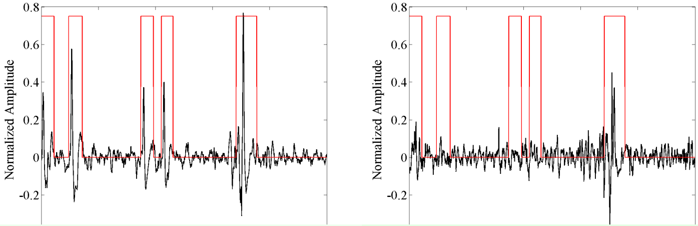
- This approach is known as CSP, introduced by K. Fukunaga and W. L. G. Koontz for classification [a],
- CSP aims at maximizing the ratio of variance between two classes, associated to windows W_i, $i = 1, 2$
- Equivalent to source separation based on nonstationarity

[a] Application of the K-L expansion to feature selection and ordering, IEEE Trans. on Computers, , no. 4, pp. 311-318, 1970

Common Spatial Pattern: Example

Epileptic focus localization, Samadi et al. [SASZJ11]

- Data : set of intracranial EEG (iEEG) of epileptic patients
- Objectives: estimate the spatial filter **B** which provides iEEG sources the most strongly related to interictal discharges (IED)
- The 2 classes are then samples in either "IED" (class 1) or "NON IED" (Class 2) time intervals
- CSP provides sources, sorted in decreasing order, according to the power in class 1 condition



III. Current Challenges and Conclusions

Part 1 Part 2

Source separation on multiple datasets

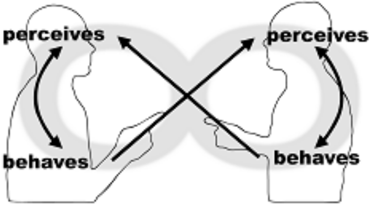
A few examples

- Group analysis: recordings of various subjects, with the same protocol
- Time evolution: recordings of the same subject at different time, with the same protocol
- Hyperscanning: recordings of 2 or more subjects at the same time

A few problems already addressed

- Group analysis: Joint BSS [LAWC09], Group ICA [CLA09]
- Hyperscanning: recordings of 2 or more subjects at the same time

Hyperscanning: Existence of brain coupling?



- Subjects in the same action-perception loop
- Perception and action share common neural networks
- Some real-life situations: gestual and spoken communication, collective sport, games, art, etc.

Joint (Blind) Source Separation

Principles of Source Separation

Compute spatial filter which

- provide sources maximally independent within the data set

Principles of Joint Source Separation

Compute a spatial filter which

- provide dependent sources, across the (2 or more) data sets
- remove independent sources, across the (2 or more) data sets

Data sets for JSS ?

- Different subjects, same experimental conditions
- Same subjects, different times or different experimental conditions

Joint (Blind) Source Separation by Joint Diagonalization

For N data sets
[CPCG12]

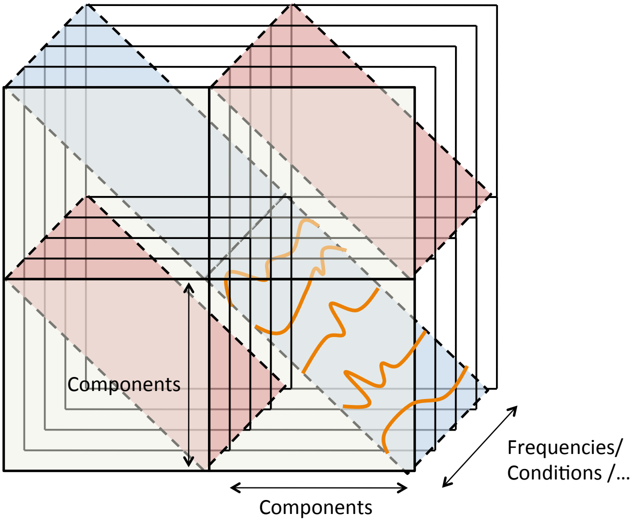
$$\begin{bmatrix} \mathbf{B}_1^T & & & \\ & \mathbf{B}_2^T & & \\ & & \ldots & \\ & & & \mathbf{B}_N^T \end{bmatrix} \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \ldots & \mathbf{X}_{1N} \\ \mathbf{X}_{21} & \mathbf{X}_{22} & & \\ \ldots & & \ldots & \\ \mathbf{X}_{N1} & & & \mathbf{X}_{NN} \end{bmatrix}_f \begin{bmatrix} \mathbf{B}_1 & & & \\ & \mathbf{B}_2 & & \\ & & \ldots & \\ & & & \mathbf{B}_N \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \ldots & \Lambda_{1N} \\ \Lambda_{21} & \Lambda_{22} & & \\ \ldots & & \ldots & \\ \Lambda_{N1} & & & \Lambda_{NN} \end{bmatrix}$$

For 2 data sets

e.g. for EEG of 2 subjects

$$\begin{cases} \mathbf{B}_1^T \ \mathbf{X}_{11}^k \ \mathbf{B}_1 = \Lambda_{11} \\ \mathbf{B}_2^T \ \mathbf{X}_{22}^k \ \mathbf{B}_2 = \Lambda_{22} \\ \mathbf{B}_1^T \ \mathbf{X}_{12}^k \ \mathbf{B}_2 = \Lambda_{12} \end{cases}$$

JSS by Joint Diagonalization of Cospectral Matrices



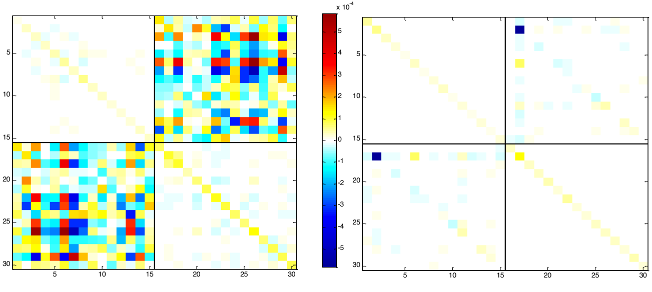
What matrices?

Steady State Visually Evoked Potentials (SSVEP)

SSVEP?

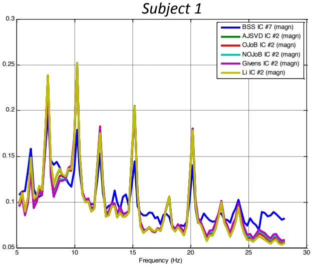
- Natural responses to visual stimulation at different frequencies: when the retina is stimulated by visual stimulus in the range $[3.5, 75]$ Hz, the brain generates electrical activity at the same (or harmonic) frequencies
- Ground truth

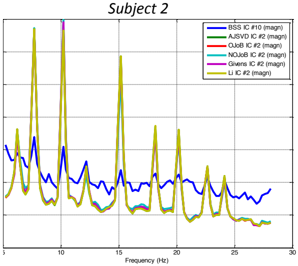
Average source cospectra in $[2, 28]$ Hz



Results with JSS: estimated sources

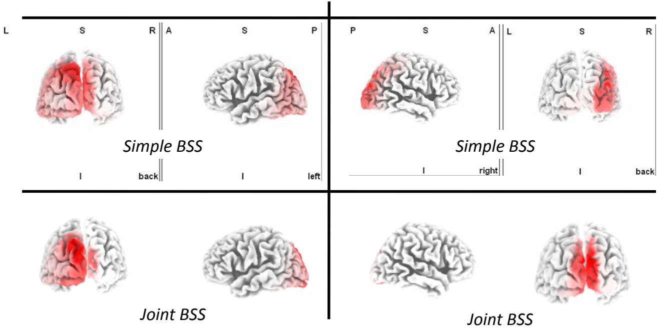
Power spectrum of the estimated source





Results with JSS: localisation with sLORETA

Localization of the estimated source



Challenges for source separation on multiple datasets

More challenging problems

- Analysis of dynamics behaviour
- Multimodal data: EEG and MEG, EEG and IRM, etc.

Source extraction

Two opposite situations

- In high-dimension data (EEG, IRM): high computational cost, only a few sources of interest
- In low-dimension data: underdetermined problem, tricky to separate all the sources

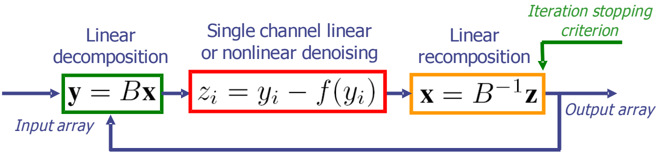
Source extraction with references: fECG extraction

Problems

- A few recordings: up to 10 electrodes
- Multidimensional sources: ECG is at least 3-D
- Larger number of sources: mother ECG, fetal ECG, EMG
- Reference is R-peak signal of mother ECG

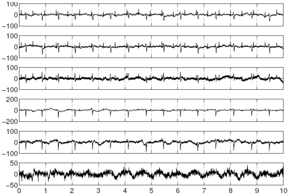
Nonlinear Deflation

[SJS10]

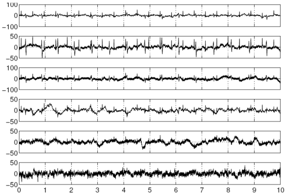


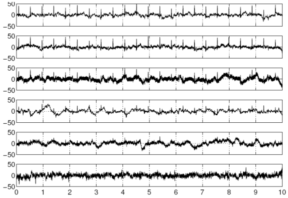
Source extraction with references: fECG extraction

1st and 2nd iterations: mother ECG is still dominant



4st and 5nd iterations: mECG disappeared, fECG appears

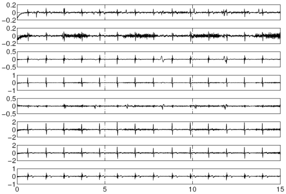


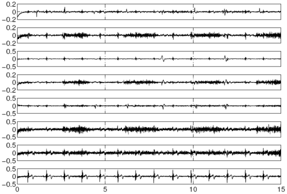


Source extraction with references: removing ECG artefact

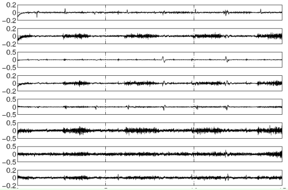
Removing ECG artefact in diaphram EMG for respiratory monitoring

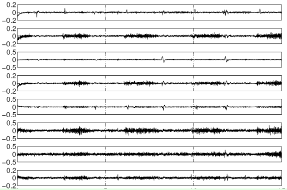
1st and 2nd iterations: ECG are still present





4st and 5nd iterations: ECG disappeared





Conclusions

Source separation

- Source separation: common problem in biomedical engineering
- Usually not blind: many priors in t or f domain, reference, etc.
- Solid theoretical framework: ICA, 2nd order methods (AJD), etc.
- Priors are useful for (1) providing more efficient methods, (2) extracting some sources and (3) labelling the estimated sources

Trends and Challenges

- Trends: Bayesian modeling; considering non negativity and sparsity
- Open issues: dynamics sources and multidimensional sources
- Challenges: extend theoretical framework for source separation and extraction in multimodal recordings

Gracias

Thank you for your attention
Questions?

Bibliography

S. Achard and C. Jutten.
Identifiability of post nonlinear mixtures.
IEEE Signal Processing Letters, 12(5):423–426, May 2005.

A. Belouchrani and K. Abed-Meraim.
Séparation aveugle au second ordre de sources corrélées.
In *GRETSI*, pages 309–312, Juan-Les-Pins, France, Septembre 1993.

M. Babaie-Zadeh, C. Jutten, and K. Nayebi.
A geometric approach for separating post nonlinear mixtures.
In *Proc. of the XI European Signal Processing Conf. (EUSIPCO 2002)*, volume II, pages 11–14, Toulouse, France, September 2002.

P. Comon and C. Jutten, editors.
Handbook of Blind Source Separation, Independent Component Analysis and Applications.
Academic Press, Oxford UK, Burlington USA, 2010.
ISBN: 978-0-12-374726-6, 19 chapters, 830 pages. hal-00460653.

V.D. Calhoun, J. Liu, and T. Adali.
A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data.
Neuroimage, 45(10):163–172, 2009.

P. Comon.
Independent Component Analysis, a new concept ?
Signal Processing, Elsevier, 36(3):287–314, April 1994.
Special issue on Higher-Order Statistics. hal-00417283.

Marco Congedo, Ronald Phlypo, and Jonas Chatel-Goldman.
Orthogonal and Non-Orthogonal Joint Blind Source Separation in the Least-Squares Sense.
In *Proceedings of EUSIPCO 2012*, pages 1885–1889, Bucharest, Roumanie, August 2012.

G. Darmois.
Analyse générale des liaisons stochastiques.
Rev. Inst. Intern. Stat., 21:2–8, 1953.

Y. Deville and S. Hosseini.
Recurrent networks for separating extractable-target nonlinear mixtures, part I: Non-blind configurations.
Signal Processing, 89(4):378–393, 2009.

J. Eriksson and V. Koivunen.
Blind identifiability of class of nonlinear instantaneous ICA models.
In *Proc. of the XI European Signal Proc. Conf. (EUSIPCO2002)*, volume 2, pages 7–10, Toulouse, France, September 2002.

R. Gribonval and S. Lesage.
A survey of sparse components analysis for blind source separation: principles, perspectives and new challenges.
In *Proc. ESANN 2006*, Bruges, Belgium, April 2006.

S. Hosseini and Y. Deville.
Blind separation of linear-quadratic mixtures of real sources using a recurrent structure.
In *Proc. of the 7th Int. Work-Conference on Artificial and Natural Neural Networks (IWANN2003), Lecture Notes in Computer Science, vol. 2686*, pages 241–248, Menorca, Spain, June 2003. Springer-Verlag.

J. Hérault, C. Jutten, and B. Ans.
Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé.
In *Actes du Xeme colloque GRETSI*, pages 1017–1022, Nice, France, 20-24 Mai 1985.

A. Hyvärinen and P. Pajunen.
Nonlinear independent component analysis: Existence and uniqueness results.
Neural Networks, 12(3):429–439, 1999.

C. Jutten, M. Babaie-Zadeh, and S. Hosseini.
Three easy ways for separating nonlinear mixtures ?
Signal Processing, 84(2):217–229, February 2004.

A. M. Kagan, Y. V. Linnik, and C. R. Rao.
Characterization Problems in Mathematical Statistics.
Probability and Mathematical Statistics. Wiley, New York, 1973.

Yi-Ou Li, T. Adali, Wei Wang, and V.D. Calhoun.
Joint blind source separation by multiset canonical correlation analysis.
Signal Processing, IEEE Transactions on, 57(10):3918 –3929, oct. 2009.

R. Llinares, J. Igual, A. Salazar, and A. Camacho.
Semi-blind source extraction of atrial activity by combining statistical and spectral features.
Digital Signal Processing, 21(2):391–403, 2011.

K. Matsuoka, M. Ohya, and M. Kawamoto.
A neural net for blind separation of nonstationary signals.
Neural Networks, 8(3):411–419, 1995.

L. Molgedey and H. Schuster.
Separation of a mixture of independent signals using time delayed correlation.
Physical Review Letters, 72:3634–3636, 1994.

D.T. Pham and J.-F. Cardoso.
Blind separation of instantaneous mixtures of nonstationary sources.
IEEE Trans. on Signal Processing, 49(9):1837–1848, 2001.

R. Phlypo, V. Zarzoso, and I. Lemahieu.
Source extraction by maximising the variance in the conditional distribution tails.
IEEE Trans on Signal Processing, 58(1):305–316, 2010.

S. Samadi, L. Amini, H. Soltanian-Zadeh, and C. Jutten.
Identification of brain regions involved in epilepsy using common spatial pattern.
In C. Richard A. Ferrari, P. Djuric, editor, *IEEE Workshop on Statistical Signal Processing (SSP2011)*, pages 829–832, Nice, France, 2011. IEEE.

R. Sameni, C. Jutten, and M. Shamsollahi.
A deflation procedure for subspace decomposition.
IEEE Transactions on Signal Processing, 58(4):2363–2374, 2010.

H.L. Nguyen Thi and C. Jutten.
Blind sources separation for convolutive mixtures.
Signal Processing, 45:209–229, 1995.

A. Taleb and C. Jutten.
On underdetermined source separation.
In *Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference - Volume 03*, pages 1445–1448, Washington, DC, USA, 1999. IEEE Computer Society.

L. Tong, V. Soon, R. Liu, and Y. Huang.
Amuse: a new blind identification algorithm.
In *Proc. ISCAS*, New Orleans, USA, 1990.

Thato Tsalaile, Reza Sameni, Saied Sanei, Christian Jutten, and Jonathon Chambers.
Sequential Blind Source Extraction For Quasi-Periodic Signals With Time-Varying Period.
IEEE Transactions on Biomedical Engineering, 56(3):646–655, 2009.

D Yellin and E. Weinstein.
Criteria for multichannel signal separation.
IEEE Trans. Signal Processing, pages 2158–2168, August 1994.