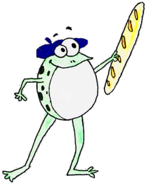
Separation of sources:
From early history, to recent advances

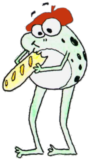
Christian Jutten

June 2024

Part 1: Early history of source separation

Christian Jutten and Pierre Comon

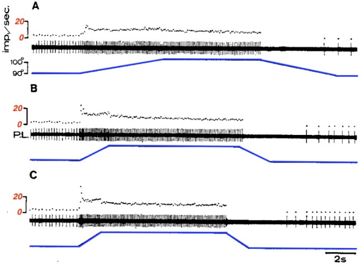




inspired by the talk presented in GRETSI 2023

The story began in 1982

- During "Neurosciences et Sciences de l'Ingénieur", Hérault, Ans and I discussed with neuroscientists about motion decoding in vertebrates.

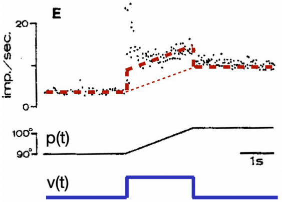


- Position and speed recorded by spindle receptors (in muscle tendons) and transmitted to the brain
- Static and dynamic fibers
- But, in each one, position and speed are mixed!!!

Question: How is the brain able to separate position and speed?

Modelling the problem

- Joint with fixed position = constant frequency
- During motion = non zero speed is superimposed to the position
- Primary fibers (f_I) are static, while secondary (f_{II}) are dynamic



- Linear model:

$$\begin{cases} f_I(t) = a_{11}p(t) + a_{12}v(t) \\ f_{II}(t) = a_{21}p(t) + a_{22}v(t) \end{cases}$$

with $a_{11} > a_{21}$ and $a_{12} < a_{22}$

- Compact notation: $\boldsymbol{x}(t) = (f_I(t), f_{II}(t))^T$ and $\boldsymbol{s}(t) = (p(t), v(t))^T$ leads to:

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t)$$

- But $\boldsymbol{A}$ and $\boldsymbol{s}(t)$ are unknown!

Knowing only $\boldsymbol{x}(t)$, is it possible to estimate $\boldsymbol{s}(t)$?

Toward a first solution

The problems

- $\boldsymbol{A}$ can be assumed to be invertible, but is unknown
- Estimate $\boldsymbol{s}(t)$ only from recordings $\boldsymbol{x}(t)$ is ill-posed

Assumptions

- $p(t)$ and $v(t)$ are statistically independent
- Strange, since $v(t) = \frac{dp}{dt}(t)$!?
- In fact, at a unique time instant t, knowledge of $p(t)$ provides no information of $v(t)$, and vice versa!

Intuition on Independent Component Analysis

- Estimate $\boldsymbol{B}$, "inverse" of the mixing $\boldsymbol{A}$, so that $\hat{\boldsymbol{s}}(t) = \boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{x}(t)$ has statistical independent components
- One can prove the solution is non unique, with scale and permutation indeterminacies.

First algorithm

Adaptive algorithm (GRETSI 1985, [HJA85])

- Due to scale indeterminacy → $\boldsymbol{B}$ with 1 on the main diagonal
- Tuning the b_{ij} : $\Delta b_{ij}(t) = \mu g(y_i(t).h(y_j(t))$, where $g(.)$ and $h(.)$ are NL odd functions,
- $\Delta b_{ij}(t) = 0$ if $E[g(y_i(t).h(y_j(t))] = 0$, i.e., y_i and y_i approx. stat. indép.

Intuitively

- $E[y_i(t).y_j(t)]$ → only decorrelation of y_i and y_j
  - Decorrelation ≠ Independance: it's only a first step
  - With this rule, $\Delta b_{ij} = \Delta b_{ji}$ which would imply a symetric $\boldsymbol{B}$: not relevant!
- $E[g(y_i(t).h(y_j(t))]$ → approx. of independance of y_i and y_j
- No proof identifiability/uniquenes and convergence (in 1985))

Context in middle of 1980's

Decorrelation or independance?

- In 1985, statistical independance is not usual:
  - With Gaussian assumptions, decorrelation is enough;
  - First workshop on HOS, Vail (Colorado) in 1989.

Boom of source separation

- J.-F. Cardoso & P. Comon (1987-...): theoretical foundations of ICA.
- Source separation & ICA: discussed in a very active working group du GdR TDSI/Isis since 1988, up to 2000
- Working Group européen ATHOS (1992-1995) managed by P.Comon
- Interest of the "neural networks" community latter, after 1995 with Bell & Sejnowski (USA), Oja & Hyvärinen (Finland), Amari & Cichocki (Japan)

Is the problem solvable?

- **Darmois'Theorem (1953) [Dar53]** Let be s_n stat. independent random variables, and

$$x_1 = \sum_n a_n s_n \quad \text{et} \quad x_2 = \sum_n b_n s_n.$$

Then, if x_1 and x_2 are stat. independent, too:

- if s_n is non Gaussian then $a_n b_n = 0$
- if $a_n b_n \neq 0$ then s_n is Gaussian

Conclusion: impossible if s_n Gaussian **AND** independent and identically distributed (iid)

When is the problem solvable?

- Darmois ⇒ s_n iid **AND** NON Gaussian (sufficient condition)
  Leads to ICA, based on high-order statistics (HOS)
  - Static mixing $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}$: if x_i pairwise independent, then $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}$, typical indeterminacies (Comon 1991)
  - **Limitation:** if $\boldsymbol{x} = \boldsymbol{A}_g\boldsymbol{s}_g + \boldsymbol{A}_h\boldsymbol{s}_h$
    with $\boldsymbol{s}_g$ Gaussian, then $\boldsymbol{A}_g$ is never identifiable (example: additive Gaussian noise, or some Gaussian sources).
- Darmois ⇒ s_n NON iid **AND** Gaussian (sufficient condition)
  Leads to second-order statistics (SOS) methods
  - identically distributed AND NON independent: colored signals (AMUSE, Tong et al., 1990; SOBI, Belouchrani et al. 1997)
  - independent AND NON identically distributed: nonstationary signals (Matsuoka et al., 1995; Pham, Cardoso, 2001)

Identifiability results

Linear "instantaneous" mixtures

- Comon in HOS 1991 and SP 1994 [Com94]
- Assumption: iid sources, mutually independent with at most one Gaussian, regular mixing matrix
- $\boldsymbol{A}$ identifiable up diagonal $\mathbf{D}$ and permutation $\mathbf{P}$ matrices $\Rightarrow$ sources with scale and permutation undeterminacies

Linear convolutive mixtures

- Yellin and Weinstein in IEEE T. on SP, 1994 [YW94]
- Assumption: sources mutually independent with condition on cross-spectra, invertible mixing matrix (with entries are LTI filters)
- $\boldsymbol{A}$ identifiable up diagonal $\mathbf{D}(\mathbf{z})$ and permutation $\mathbf{P}$ matrices $\Rightarrow$ sources with unknow filter and permutation undeterminacies

Nonlinear mixtures See Part II

Other approaches and priors

Source separation in transformed domains

- Consider a mapping $\mathcal{T}$ which preserves linearity (e.g., Fourier transform, DCT, wavelet transform...)
- $\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t)$ becomes:

$$\mathcal{T}(\boldsymbol{x}(t)) = \boldsymbol{A}\mathcal{T}(\boldsymbol{s}(t) \quad (1)$$

$$\boldsymbol{x}(\nu) = \boldsymbol{A}\boldsymbol{s}(\nu) \quad (2)$$

- Solving in the transformed domain (can be simpler)
- Then, use $\mathcal{T}^{-\infty}$ for coming back in the initial domain

Results based on other priors

- Bounded sources
- Discrete-valued sources (e.g., in communications)
- Nonnegative sources and mixtures (in time or spectral domains)
- Sparsity... key to solve underdetermined source separation

Many algorithms and applications

First algorithms

- CoM (Contrast Maximization): Comon 91 [Com92]
- Joint diagonalization: JADE (1993), [CS93]
- AMUSE (1990), SOBI (93), etc. [BAM93]
- Equivariant algotithm (Cardoso, Laheld, 1994) [CL96]
- Infomax (Bell, Sejnowski, 1995) [BS95], FastIca (Hyvärinen, Oja, 1999) [Hyv99]

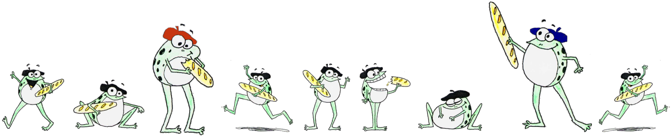
Applications

- Speech and music enhancement and separation
- Biomedical engineering: ECG, EEG, EMG, ...
- Hyperspectral imaging
- Chemistry and physics

Why such a success?

- ICA much powerfull than PCA, and with physical meaning
- mixing models (even very simple) relevant for various applications
- rigourous theoretical foundations
- efficient algorithms, with proof of convergence

And above all, an important swarm in the Gdr ISIS!



French GdR ISIS plays an essential role

Working Group: HOS and source separation

- From 1988 to 2004: 3 to 4 1-day meetings per year
- Managed by J.-F. Cardoso and then E. Moreau
- Friendly meetings; tutorial, PhD talks, discussions
- Strong interactions: academics, industrials and PhD students
- About 35 PhD on BSS in France defended from 1991 to 2004

GdR ISIS funded collaborative projects

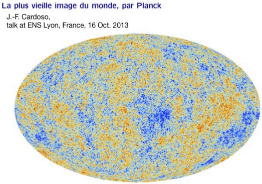
- With support of "Club of industrial parters"
- E.g., in 2000, Févotte, Vincent and Gribonval received grant for developping international competition for speech/music source separation...

BSS Success Stories

Many awards

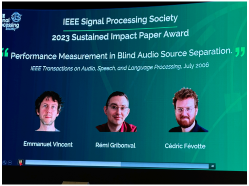
- Many best paper and IEEE SPS awards, 2 CNRS Silver medals
- Some European Research Council grants
- Front page of Washington Post in 2014

Jean-François Cardoso
CNRS Silver Medal
in 2014



Front page of Washington Post in 2014

Vincent, Févotte and Gribonval in the spotlight at ICASSP 2024, in Seoul



Pierre Comon
CNRS Silver Medal
in 2018

Part 2: Advances in nonlinear source separation

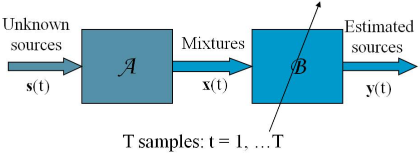
Christian Jutten with
Massoud Babaie-Zadeh, Leonardo Duarte, Bahram Ehsandoust, Bertrand Rivet and Anisse Taleb

Model and question

Model

- K noiseless nonlinear (NL) mixtures of P independent sources

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



Question

- Assuming the NL mixing mapping $\mathcal{A}$ is invertible, is it possible to estimate an inverse mapping $\mathcal{B}$ using independence ?
- In other words: output independence $\Leftrightarrow$ s source separation ?

Darmois's results [Dar53]

Undeterminacy

- Let s_i and s_j be 2 independent random variables, $f_i(s_i)$ and $f_j(s_j)$ are independent too
- Source separation is achieved if $y_i = h_i(s_j)$, i.e. the global mapping $\mathcal{G} = \mathcal{B} \circ \mathcal{A}$ is diagonal, up to a permutation.
- Such diagonal (up to a permutation) mappings $\mathcal{G}$ will be defined as *trivial mappings*

Nonlinear mixtures are non identifiable using ICA

- It always exists non diagonal nonlinear mixing mappings which preserve independence
- Darmois (1953) proposed a general method for constructing such mappings. The idea has then been used by Hyvärinen and Pajunen (1999 [HP99]).

A simple example

Consider 2 independent Gaussian variables x_1 and x_2 with joint pdf

$$p_{\mathbf{x}}(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)$$

Consider the following mapping and its Jacobian

$$\begin{cases} x_1 & = & r\cos\theta \\ x_2 & = & r\sin\theta \end{cases} \qquad J = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

The joint pdf of r and θ is

$$p_{r\theta}(r, \theta) = \begin{cases} \frac{1}{2\pi} r \exp(-r^2) & \text{if } (r, \theta) \in \mathbb{R}^+ \times [0, 2\pi] \\ 0 & \text{otherwise} \end{cases}$$

Althought r and θ are depending both of x_1 and x_2, they are statistically independent!

Conclusions

General results

- Statistical independence is not sufficient for insuring identifiability of NL mixtures
- For any sources, it exists invertible mappings with non diagonal Jacobian (i.e. mixing or nontrivial mappings) which preserve statistical independence $\rightarrow$ generally, ICA not efficient
- If the mapping can be identified, source can be recovered up to a NL mapping (and permutation)

For overcoming the problem,

- Approaches reducing the set of nontrivial mappings preserving independence
- Use additional priors, e.g., sparsity, non iid sources, Gaussian processes...

Structural constraints: general results (1/3)

All these idea are detailed in the nice Taleb's paper of IEEE Trans. on SP [Tal02].

Trivial mappings: definition

- Definition: $\mathcal{H}$ is a trivial mapping if it transforms *any* random vector with independent components in another random vector with independent components.
- The set of the trivial mappings will be denoted Z

Trivial mappings: properties

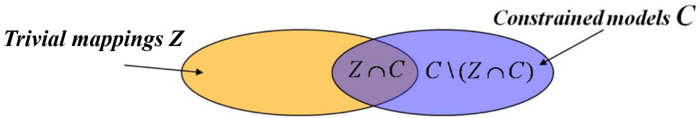
- A trivial mapping is then a mapping preserving independence for *any* random vector
- It can be shown that a trivial mapping satisfies $H_i(s_1, \ldots, s_P) = h_i(s_{\sigma(i)}), \forall i = 1, \ldots, K$
- The Jacobian matrix of a trivial mapping is a diagonal matrix up to a permutation

Structural constraints: general results (2/3)

There is an infinity of nontrivial mappings preserving independence

Constrained model of mixtures

- If the mapping $\mathcal{G} = \mathcal{B} \circ \mathcal{A}$ is constrained in the set $\mathcal{C}$, undeterminacies can be reduced, and hopefully cancelled
- Consider $\Omega = \{F_{s_1}, \ldots, F_{s_P}\}$, the set of signal distributions such that $\exists \mathcal{G} \in \mathcal{C} - \mathcal{Z}$ (i.e., a non trivial mapping) which preserves independence for any $\omega \in \Omega$
- Ω then contains all the (particular) source distributions which cannot be separated by mapping belonging to $\mathcal{C}$.

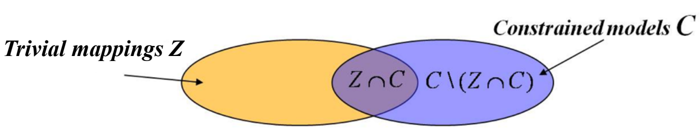


- Separation is then possible (1) for source distributions which do not belong to Ω, (2) with indeterminacies in $\mathcal{G} \in \mathcal{Z} \cap \mathcal{C}$

Structural constraints: general results (3/3)

Example of linear memoryless regular mappings

- $\mathcal{C}$ is the set of square regular matrices
- $\mathcal{Z} \cap \mathcal{C}$ is the set of square matrices which are the product of a diagonal matrix and a permutation matrix
- Ω is the set of distributions which contain at least 2 Gaussian



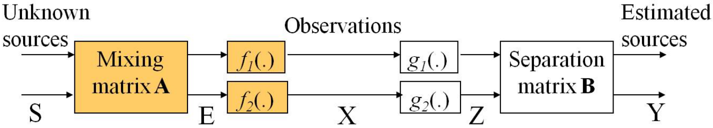
Conclusions

- For linear memoryless mixtures, source separation is possible using ICA (1) for sources which are not in Ω (i.e. at most one Gaussian) and (2) with scale and permutation undeterminacies.

Structural constraints: PNL mixtures

Post-nonlinear (PNL) mixtures

- PNL are particular nonlinear mixtures, which structural constraints : linear part, following by NL componentwise mappings.
- PNL are realistic enough : linear channel, nonlinear sensors

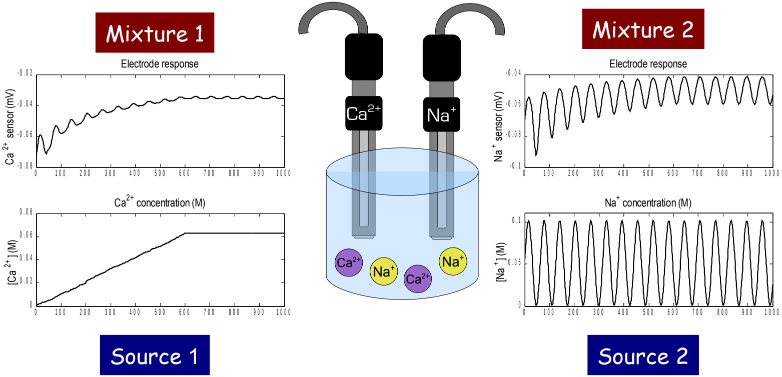


PNL identifiability (Taleb et al. 99, Achard et al. 05) with suited $\mathcal{B}$

- if (1) at most one source is Gaussian, (2) the mixing matrix has at least two nonzero entries per row and per column, and (3) the NL mappings f_i are invertible and satisfy $f_i'(0) \neq 0$, then $\boldsymbol{y}$ is independent iff $g_i \circ f_i$ is linear and $\boldsymbol{BA} = \boldsymbol{DP}$

Ion-selective sensors: the interference problem

- Aim: to estimate concentrations of several ions in a solution.
- Problem: the ion sensors are not very selective!



Solving the interference problem

Summary

- Based on the Nicolski-Eisenman model
- Method based on source silences (Duarte et al., Eusipco 2008 [DJ08])
- Bayesian approach (Duarte et al., ICA 2009 [DJM09])

These works were done by L. Duarte during his PhD thesis in GIPSA-lab (2006-2009)

The Nicolsky-Eisenman model (1/2)

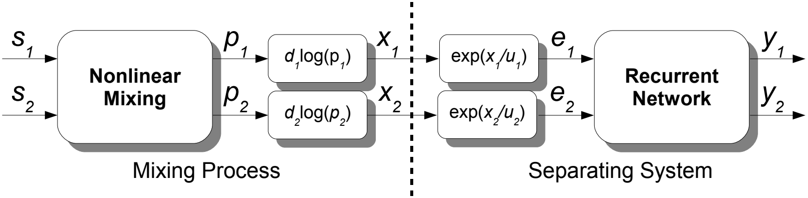
$$x_i(t) = c_i + d_i \log\left(s_i(t) + \sum_{j, j \neq i} a_{ij} s_j(t)^{\frac{z_i}{z_j}}\right), \quad (3)$$

- $s_i(t) \Rightarrow$ target ion concentration; $s_j(t) \Rightarrow$ interfering ions concentrations
- c_i, d_i, $a_{ij} \Rightarrow$ mixing model parameters;
- z_i and $z_j \Rightarrow$ valences of the ions i and j
- When $z_i = z_j \Rightarrow$ Post-nonlinear (PNL) mixing model.

Additionnal conditions

- *We are interested in the case in which* $z_i \neq z_j$.
- *We consider a scenario with two ions and two electrodes*.

The Nicolsky-Eisenman model (2/2)



Resulting model for 2 sensors and 2 ions

$$\begin{aligned} x_1(t) &= d_1 \log\left(s_1(t) + a_{12} s_2(t)^k\right) \\ x_2(t) &= d_2 \log\left(s_2(t) + a_{21} s_1(t)^{\frac{1}{k}}\right) \end{aligned}, \tag{4}$$

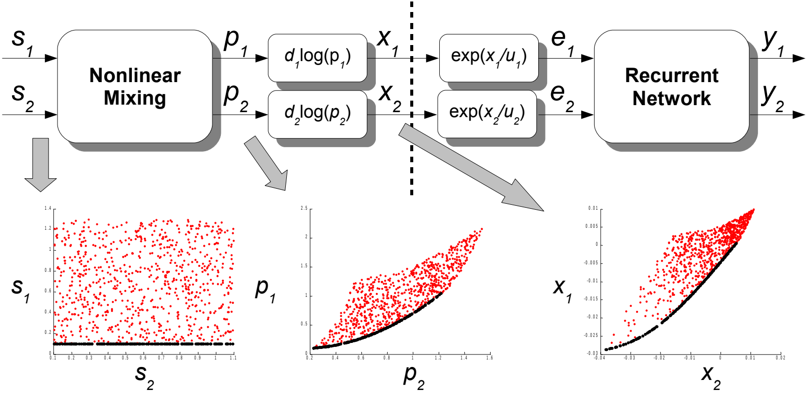
with $k = z_1/z_2$.

Assumptions

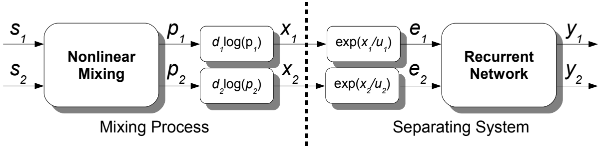
1. The sources are statistically independent;
2. The sources are positive and bounded, i.e., $s_i(t) \in [S_i^{min}, S_i^{max}]$, where $S_i^{max} > S_i^{min} > 0$;
3. The mixing system is invertible in the region given by $[S_1^{min}, S_1^{max}] \times [S_2^{min}, S_2^{max}]$;
4. k (the ratio between the valences) is known and takes only positive integer values;

Using the prior: one source is silent

- Additional assumption: during some periods of time, the concentration of one ion is constant (zero-variance).



Basic idea in equations



- During the silent periods ($s_1(t) = S_1 = cte$):

$$\begin{aligned} p_1(t) &= S_1 + a_{12} s_2(t)^k \\ p_2(t) &= s_2(t) + a_{21} S_1^{\frac{1}{k}} \end{aligned} \quad . \tag{5}$$

- In the (p_1, p_2) plane, we have a polynomial of order k:

$$p_1(t) = S_1 + a_{12}(p_2(t) - a_{21} S_1^{\frac{1}{k}})^k. \tag{6}$$

$$p_1(t) = \sum_{i=0}^{k} \varphi_i p_2(t)^i, \tag{7}$$

- Idea: e_1 must be a polynomial of order k, too

How to detect the silent periods?

- During the silent periods of $s_1(t)$,

$$\begin{array}{rcl} x_1 & = & g_1(s_2) \\ x_2 & = & g_2(s_2) \end{array} . \tag{8}$$

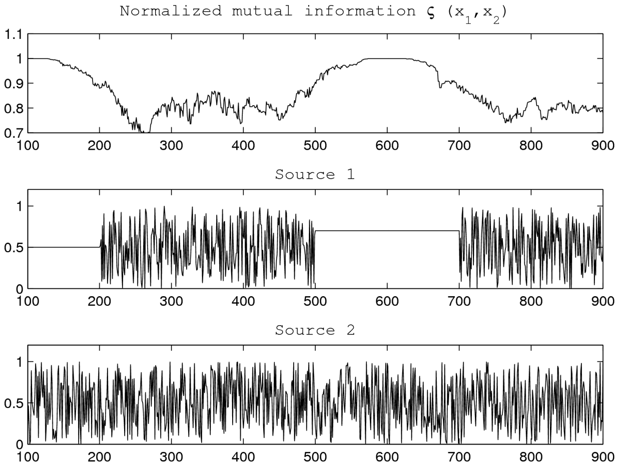
$\rightarrow$ maximum (nonlinear) correlation between x_1 and x_2

- Normalized mutual information

$$\varsigma(x_1, x_2) = \sqrt{1 - \exp(-2I(x_1, x_2))} \tag{9}$$

- $\varsigma(x_1, x_2) = 0$ when x_1 and x_2 are statistically independent;
- $\varsigma(x_1, x_2) \rightarrow 1$ when there is a deterministic relation between x_1 and x_2;

Result for silent periods detection



Motivations for using the Bayesian approach

- Prior information is available

$$x_{it} = e_i + d_i \log_{10}\left(s_{it} + \sum_{j, j\neq i} a_{ij} s_{jt}^{\frac{z_i}{z_j}} \right) + n_{it}, \qquad (10)$$

  - e_i takes value in the interval $[0.05, 0.35]$; (Gruen 2007)
  - Theoretical value for the Nerstian slope $\Rightarrow d_i = RT \ln(10)/z_i F$ ($0.059V$ for room temperature); (Fabri et al, 2003)
  - Always non-negative. Very often in the interval $[0, 1]$;
  - The sources are positive.

- Takes noise into account;
- In contrast to ICA, the statistical independence is rather a working assumption in the Bayesian approach (Fevotte et al. 2006);
- May work even if the number of samples is small.

Bayesian source separation method: problem and notations

- Problem: given $\mathbf{X}$, estimate the unknown parameters $\boldsymbol{\theta} = [\mathbf{S}, \mathbf{A}, \mathbf{d}, \mathbf{e}, \boldsymbol{\sigma}, \phi]$;
- $\mathbf{S} \Rightarrow$ sources;
- $\phi \Rightarrow$ sources hyperparameters;
- $\mathbf{A} \Rightarrow$ selectivity coefficients;
- $\mathbf{d} \Rightarrow$ Nerstian slopes;
- $\mathbf{e} \Rightarrow$ offset parameters;
- $\boldsymbol{\sigma} \Rightarrow$ noise variances.

Bayesian source separation method: an overview

- Problem: given $\mathbf{X}$, estimate the unknown parameters $\boldsymbol{\theta} = [\mathbf{S}, \mathbf{A}, \mathbf{d}, \mathbf{e}, \boldsymbol{\sigma}, \phi]$;
- In the Bayesian approach, estimation of $\boldsymbol{\theta}$ is based on the posterior information

$$p(\boldsymbol{\theta}|\mathbf{X}) \propto p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{11}$$

- The likelihood function is given by:

$$p(\mathbf{X}|\boldsymbol{\theta}) = \prod_{t=1}^{n_d}\prod_{i=1}^{n_c}\mathcal{N}_{x_{it}}\left(e_i + d_i \log\left(\sum_{j=1}^{n_s} a_{ij} s_{jt}^{z_i/z_j}\right), \sigma_i^2\right), \tag{12}$$

assuming an additive i.i.d. Gaussian noise vector which is spatially independent.

Prior definitions

- Log-normal prior distribution for the sources (non-negative distribution)

$$p(s_{jt}) = \frac{1}{s_{jt}\sqrt{2\pi\sigma_{s_j}^2}} \exp\left(-\frac{(\log(s_{jt}) - \mu_{s_j})^2}{2\sigma_{s_j}^2}\right) \mathbb{1}_{[0,+\infty[}(s_{jt}), \tag{13}$$

- Motivations
  - The estimation of $\phi_j = [\mu_{s_j}\ \sigma_{s_j}^2]$ is not difficult, since we can define a conjugate pair.
  - Ionic activities are expected to have a small variation in the logarithmic scale.
- The sources are assumed i.i.d. and statistically mutually independent:

$$p(\mathbf{S}) = \prod_{j=1}^{n_s}\prod_{t=1}^{n_d} p(s_{jt}), \tag{14}$$

Prior definitions (cont.)

- Sources parameters $\phi_j = [\mu_{s_j}\ \sigma^2_{s_j}]$

$$p(\mu_{s_j}) = \mathcal{N}(\tilde{\mu}_{s_j}, \tilde{\sigma}^2_{s_j}), \quad p(1/\sigma^2_{s_j}) = \mathcal{G}(\alpha_{\sigma_{s_j}}, \beta_{\sigma_{s_j}}) \tag{15}$$

- Selectivity coefficients a_{ij}: very often within $[0, 1]$

$$p(a_{ij}) = \mathcal{U}(0, 1) \tag{16}$$

- Nernstian slopes d_i: ideally $0.059V$ at room temperature

$$p(d_i) = \mathcal{N}(\mu_{d_i} = 0.059/z_i, \sigma^2_{d_i}) \tag{17}$$

- Offset parameters e_i lie in the interval $[0.050, 0.350]V$

$$p(e_i) = \mathcal{N}(\mu_{e_i} = 0.20, \sigma^2_{e_i}) \tag{18}$$

- Noise variances σ_i:

$$p(1/\sigma^2_i) = \mathcal{G}(\alpha_{\sigma_i}, \beta_{\sigma_i}) \tag{19}$$

The posterior distribution

- The posterior distribution is given by

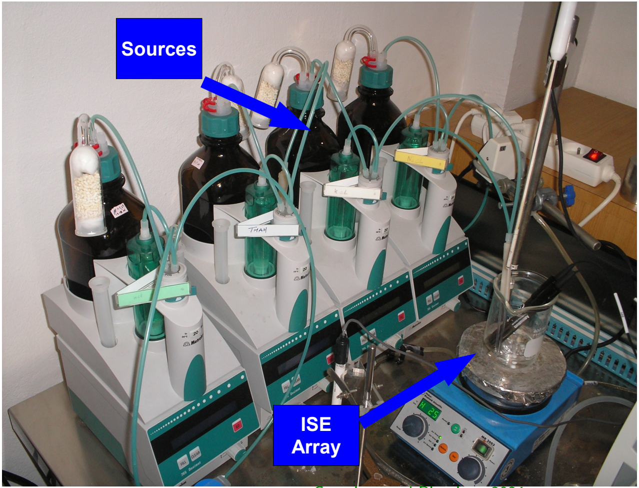
$$p(\boldsymbol{\theta}|\mathbf{X}) \propto p(\mathbf{X}|\boldsymbol{\theta}) \cdot \prod_{j=1}^{n_s}\prod_{t=1}^{n_d} p(s_{jt}|\mu_{s_j}, \sigma_{s_j}^2) \cdot \prod_{j=1}^{n_s} p(\mu_{s_j}) \cdot \prod_{j=1}^{n_s} p(\sigma_{s_j}) \cdot \prod_{i=1}^{n_c}\prod_{j=1}^{n_s} p(a_{ij}) \cdot \prod_{i=1}^{n_c} p(e_i) \cdot \prod_{i=1}^{n_c} p(d_i) \cdot \prod_{i=1}^{n_c} p(\sigma_i) \tag{20}$$

- Bayesian MMSE estimator $\Rightarrow \boldsymbol{\theta}_{MMSE} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{X}) d\boldsymbol{\theta}$
  (Difficult to calculate!)
- Given $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^M$ (samples drawn from $p(\boldsymbol{\theta}|\mathbf{X})$), the Bayesian MMSE estimator can be approximated by:

$$\widetilde{\boldsymbol{\theta}}_{MMSE} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{\theta}^i \tag{21}$$

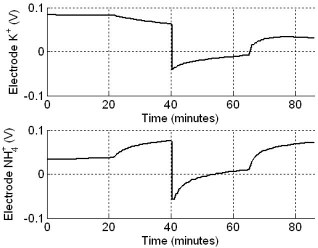
Results on real data

- ISE array ($NH_4^+ - ISE$ and $K^+ - ISE$)

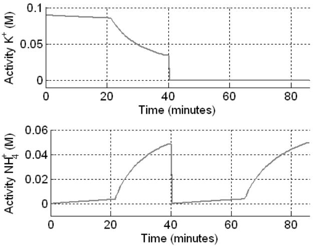


Results on real data (cont.)

$n_d = 169$, $SIR_1 = 25.1\,\mathrm{dB}$, $SIR_2 = 23.7\,\mathrm{dB}$, $SIR = 24.4\,\mathrm{dB}$



(a) ISE array response.

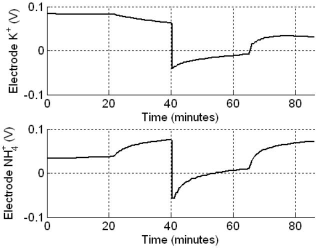


(b) Actual sources.

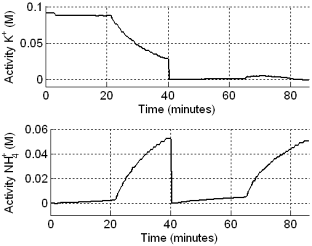
- Since the sources are clearly dependent here, an ICA-based method failed in this case.

Results on real data (cont.)

$n_d = 169$, $SIR_1 = 25.1\,\mathrm{dB}$, $SIR_2 = 23.7\,\mathrm{dB}$, $SIR = 24.4\,\mathrm{dB}$



(a) ISE array response.



(b) Retrieved signals.

- Since the sources are clearly dependent here, an ICA-based method failed in this case.

Conclusions on ion concentration estimations

Based on silence

- Silence = kind of sparsity
- Main limitations:
  - Number of samples in real applications may be small.
  - Many priors haven't been used
  - The independence assumption may be rather strong, especially if a regulatory process between ions exists.

Bayesian approach

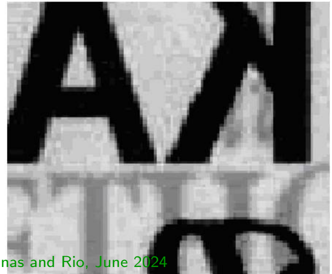
- A Bayesian nonlinear source separation is a flexible approach for processing the outputs of an ion selective electrode array;
- Good results are achieved even in tricky situations: (1) dependent sources and (2) reduced number of samples.

Show-through effect

Work developped by F. Merrikh-Bayat and M. Babaie-Zadeh, Sharif Univ. of Technology (Merrikh-Bayat et al. [MBBZJ08, MBBZJ11])

What is show-through?

- Show-through, due to paper transparency and thickness,
- Pigment oil penetration,
- Vehicle oil component, due to loss of opacity,



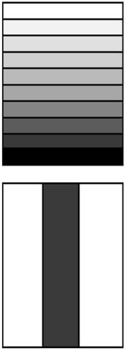
State-of-the-art

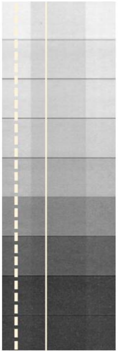
- Often applied for texts and handwritting documents: 1-side methods or 2-side methods,
- ICA assuming
  - Linear model of mixtures (Tonazzini et al., 2007 ; Ophir, Malah, 2007)
  - Nonlinear model of mixtures (Almeida, 2005 ; Sharma, 2001)
- In this work, we consider:
  - modelisation of the nonlinear mixture,
  - blurring effect.

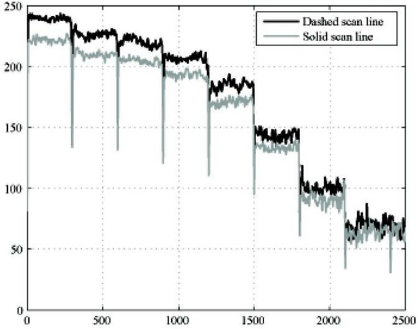
Nonlinearity of show-through: experimental study

Evidence

- Sum of luminance is NL.
- Whiter the pixel, more important is show-through. More black than black is impossible !







Nonlinearity of show-through: mathematical model

Basic equation

- Show-through has a gain which depends of the grayscale of the front image
- It leads to the model of mixtures:

$$\begin{cases} f_r^s(x,y) & = & a_1 f_r^i(x,y) + b_1 f_v^i(x,y) g_1[f_r^i(x,y)] \\ f_v^s(x,y) & = & a_2 f_v^i(x,y) + b_2 f_r^i(x,y) g_2[f_v^i(x,y)] \end{cases}$$

where i = initial, s = scanned, r = recto, v = verso, a_i and b_i denote unknown mixing parameters, and $g_i(.)$ denote nonlinear gains

Nonlinearity of show-through: mathematical model

The gain function is experimentally estimated by computing [MBBZJ08]

$$\left\{\begin{array}{lcl} g_1[f_r^i(x,y)] & = & [f_v^s(x,y) - a_1 f_r^i(x,y)]/b_1 f_v^i(x,y) \\ g_2[f_v^i(x,y)] & = & [f_v^s(x,y) - a_2 f_v^i(x,y)]/b_2 f_r^i(x,y) \end{array}\right.$$

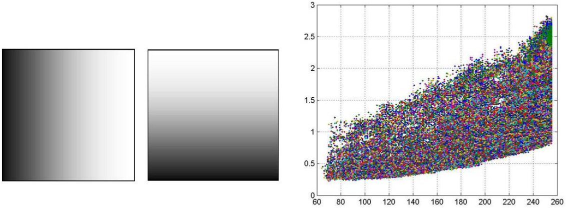


Figure: Left: face and back sides of the printed sheet used in the experiment. Right:plot of the right side of the above equation vs. $f_v^i(x,y)$ or $f_r^i(x,y)$, for each pixel.

Nonlinearity of show-through: mathematical model

Approximation of the gain function

- The gain function can be estimated by an exponential:

$$\begin{cases} g_1[f_r^i(x,y)] & = & \gamma_1 \exp[\beta_1 f_r^i(x,y)] \approx \gamma_1(1+\beta_1 f_r^i(x,y)) \\ g_2[f_v^i(x,y)] & = & \gamma_2 \exp[\beta_2 f_v^i(x,y)] \approx \gamma_2(1+\beta_2 f_v^i(x,y)) \end{cases}$$

- It leads to the approximated mixing model:

$$\begin{cases} f_r^s(x,y) & = & a_1 f_r^i(x,y) + b_1' f_v^i(x,y)[1+\beta_1 f_r^i(x,y)] \\ f_v^s(x,y) & = & a_2 f_v^i(x,y) + b_2' f_r^i(x,y)[1+\beta_2 f_v^i(x,y)] \end{cases}$$

- And finally to the bilinear model:

$$\begin{cases} f_r^s(x,y) & = & a_1 f_r^i(x,y) - l_1 f_v^i(x,y) - q_1 f_v^i(x,y) f_r^i(x,y)] \\ f_v^s(x,y) & = & a_2 f_v^i(x,y) + l_2 f_r^i(x,y) - q_2 f_r^i(x,y) f_v^i(x,y)] \end{cases}$$

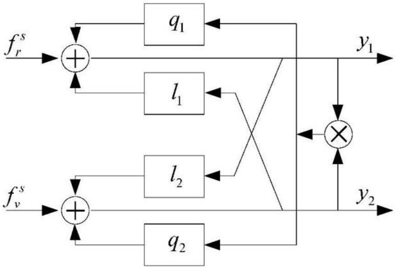
Separation structure

Recursive structure

- Studied by Deville and Hosseini ([HD03, DH09])

$$\begin{cases} f_r^s(x,y) &= a_1 f_r^i(x,y) - l_1 f_v^i(x,y) - q_1 f_v^i(x,y) f_r^i(x,y)] \\ f_v^s(x,y) &= a_2 f_v^i(x,y) + l_2 f_r^i(x,y) - q_2 f_r^i(x,y) f_v^i(x,y)] \end{cases}$$

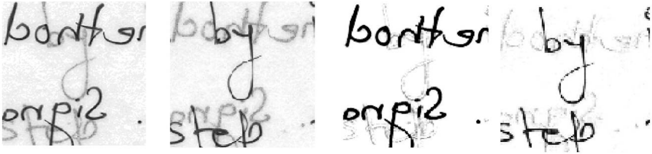
- who proposed the following recursive architecture suited to the model: one equilibrium state is the solution



Cancellation of show-through: preliminary results

Preliminary results with NL model

- Bilinear model neither always invertible, nor always stable.
- Parameters estimated by ML.



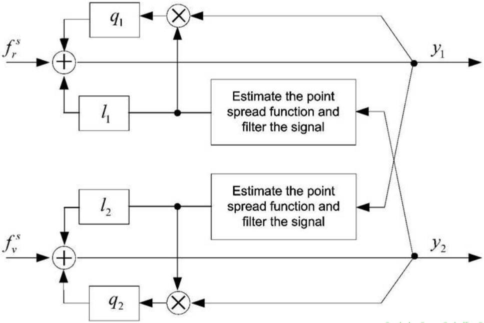
Comments

- The other side image never perfectly removed, especially when no superimposition!
- It means difference between verso image and recto image is not a simple gain
- Diffusion in the paper ⇒ blurring effect, modelled by 2D filter.

Improved model and recursive structure

Model with filtering

- The mixture is not the nonlinear superimposition of the recto (verso, resp.) image with the verso image, but with a *filtered* version of the verso (recto, resp.) image, hence the final separation structure



Cancellation of show-through

Final results with NL modeling and filtering

(a) (b) (c)

Experimental results. (a): recorded front side image. (b) and (c): estimated cleaned frontside images without (b) and with (c) filter.

Removing Show-Through in Scanned images: conclusions

Summary

- Show-through is a NL phenomenon which can be modeled by bilinear mixtures.
- In addition, the blurring effect can be modelled by a 2-D filter.
- Experimental results show the mixture model has to take into account both NL and convolutive effects.
- Other priors, like positivity of images, and of the coefficients could be exploited, e.g. by NMF or Bayesian approaches.

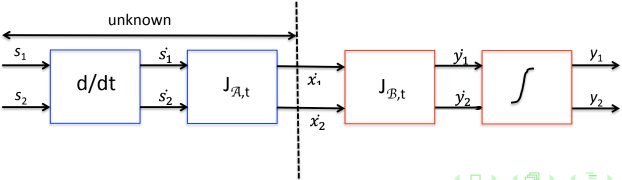
Nonlinear model: Local linear approximation (1/2)

A simple idea (Ehsandoust et al., LVA-ICA 2015 and IEEE T. SP 2016 [EBZRJ17])

- inspired by Levin's paper (2010)
- Deriving the nonlinear (time-invariant) mixture leads to

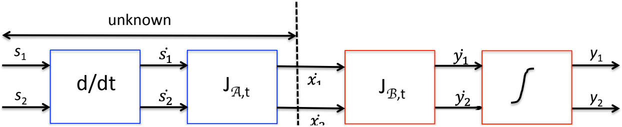
$$\boldsymbol{x}(t) = \mathcal{A}(\boldsymbol{s}(t)) \rightarrow \dot{\boldsymbol{x}}(t) = \boldsymbol{J}_{\mathcal{A},t}\dot{\boldsymbol{s}}(t) \tag{22}$$

- i.e. a linear time-varying mixtures, due to the Jacobian matrix $\boldsymbol{J}_{\mathcal{A},t}$.



Nonlinear model: Local linear approximation (2/2)

$$\boldsymbol{x}(t) = \mathcal{A}(\boldsymbol{s}(t)) \rightarrow \dot{\boldsymbol{x}}(t) = \boldsymbol{J}_{\mathcal{A},t}\dot{\boldsymbol{s}}(t)$$



Separability

- Separability of linear mixtures, up to scale and permut + cte
- Require statistical independence of $\dot{s}_r$, $r = 1, \ldots R$

Algorithm

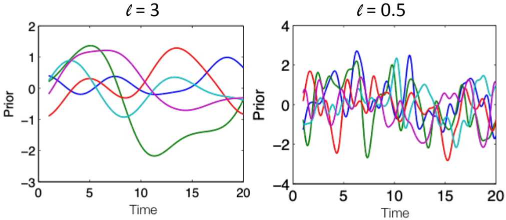
- Since $\boldsymbol{J}_{f,t}$ is linear time-varying
- convergence requires slowly varying sources
- adaptive separation algorithm, inspired by EASY (Laheld, Cardoso, 1996)

Nonlinear mixture of Gaussian process sources (1/2)

Sources as Gaussian processes

- Source separation in linear mixture can be achieved by considering relation between successive samples
- We propose to consider sources as Gaussian Processes, i.e. $s_r(t) \sim GP(m(t); k(t, t'))$
- GP are very flexible for modeling large range of colored signals

$$e^{\left(\frac{-(t-t')^2}{2\times l^2}\right)}$$



Nonlinear mixture of Gaussian process sources (2/2)

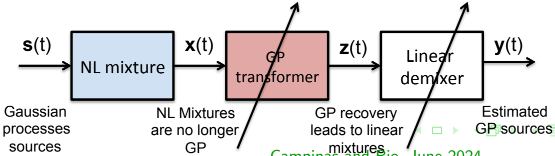
GP property of sources is lost when mapped with NL polynomials.

Theorem (Ehsandoust et al., ICASSP 2017 [ERBZJ17])

Let R sources, $s_1, \ldots, s_R$, be jointly Gaussian processes mixed by an invertible polynomial $\boldsymbol{P}$, i.e. $\boldsymbol{y} = \boldsymbol{P}(\boldsymbol{s})$. The mapping $\boldsymbol{y}$ is jointly Gaussian distributed iff $\boldsymbol{y} = \boldsymbol{P}(\boldsymbol{s}) = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{c}$, where $\boldsymbol{A}$ is a $R \times R$ matrix and $\boldsymbol{c}$ is a $R \times 1$ vector with scalar entries.

Source separation in 2 steps (Ehsandoust et al. [ERBZJ17])

- Recovering GP property cancels the nonlinear part of the mapping
- A simple linear demixer can then estimate the GP sources.



Take home message

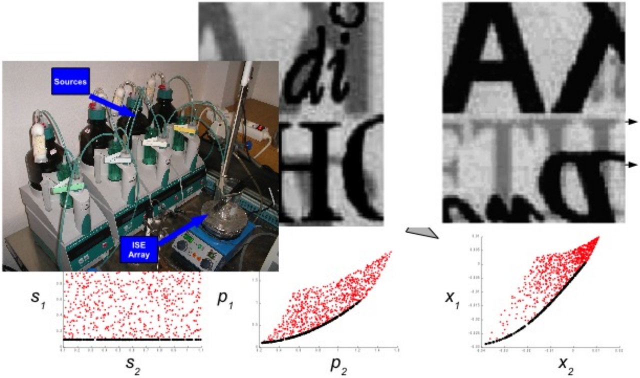
Conclusions

- Independence is not sufficient for insuring identifiability and separability in *general nonlinear mixtures*
- Independence ⇒ identifibility and separability in *constrainted NL mixtures*, e.g., PNL, bilinear, linear-quadratic models
- Priors on sources, e.g. bounded, sparse, non-negative or colored sources, can (1) provide simpler separation criterion, and (2) reduce solution indeterminacies.
- Many problems require nonlinear models: chemical sensor, scanned image processing, hyperspectral imaging, ...

News ideas to be further investigated [Ehs17]

- Replacing NL invariant model by a linear variant model
- Considering non iid sources, e.g., with Gaussian processes

Thanks for your attention!



Thanks
to the European Research Council for the project ERC-AdG CHESS,
and to Campinas University for the lecturer invitation.

Bibliography

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.

A. J. Bell and T. J. Sejnowski.
An information-maximization approach to blind separation and blind deconvolution.
Neural Computation, 7(6):1129–1159, November 1995.

J. F. Cardoso and B. Laheld.
Equivariant adaptive source separation.
IEEE Trans. on Sig. Proc., 44(12):3017–3030, December 1996.

P. Comon.
Independent Component Analysis.
In J-L. Lacoume, editor, *Higher Order Statistics*, pages 29–38. Elsevier, Amsterdam, London, 1992.

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.

J. F. Cardoso and A. Souloumiac.
Blind beamforming for non-Gaussian signals.
IEE Proceedings - Part F, 140(6):362–370, December 1993.
Special issue on Applications of High-Order Statistics.

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.

L. Duarte and C. Jutten.
A nonlinear source separation approach for the nicolsky-eisenman model.
In *CD-Rom Proceedings of EUSIPCO 2008*, Lausanne, Switzerland, sep 2008.

Leonardo Tomazeli Duarte, Christian Jutten, and Saïd Moussaoui.
Ion-Selective Electrode Array Based on a Bayesian Nonlinear Source Separation Method.
In *ICA 2009 - 8th International Conference on Independent Component Analysis and Signal Separation*, volume 5441, pages 662–669, Paraty, Brazil, March 2009. Springer.

Bahram Ehsandoust, Massoud Babaie-Zadeh, Bertrand Rivet, and Christian Jutten.
Blind source separation in nonlinear mixtures: Separability and a basic algorithm.
IEEE Transactions on Signal Processing, 65(16):4339–4352, 2017.

B. Ehsandoust.
Blind Source Separation in Nonlinear Mixtures.
PhD thesis, Univ. Grenoble Alpes (France) and Sharif Univ. of Technology (Iran), 2017.

Bahram Ehsandoust, Bertrand Rivet, Massoud Babaie-Zadeh, and Christian Jutten.
Blind compensation of polynomial mixtures of gaussian signals with application in nonlinear blind source separation.
In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4681–4685, 2017.

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.

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

F. Merrikh-Bayat, M. Babaie-Zadeh, and C. Jutten.
A nonlinear blind source separation solution for removing the show-through effect in the scanned documents.
In *Proc. of the 16th European Signal Processing Conf. (EUSIPCO2008)*, Lausanne, Switzerland, August 2008.

F. Merrikh-Bayat, M. Babaie-Zadeh, and C. Juttenl.
Linear-quadratic blind source separating structure for removing show-through in scanned documents.
International Journal of Document Analysis and Recognition, page 15 pages, 2011.
published online at Oct. 28, 2010; to appear.

A. Taleb.
A generic framework for blind source separation in structured nonlinear models.
IEEE Transactions on Signal Processing, 50(8):1819–1830, 2002.

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