Blind Source Separation (BSS) and Independent Componen Analysis (ICA)

Massoud BABAIE-ZADEH

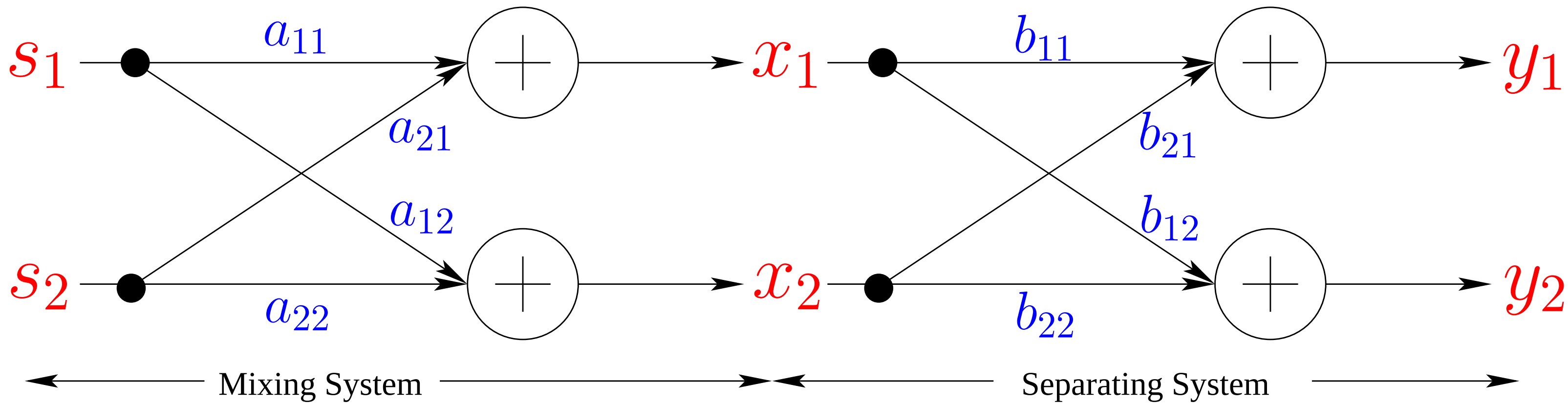
Outline

- Part I Introduction
- Part II Separability
- Part III Some famous approaches for solving BSS problem
- Part IV Extensions to ICA
- Part V Applications, my works and perspectives

Part I

Introduction to BSS and ICA

Part I Blind Source Separation (BSS)



- s_i: Original source (assumed to be independent).
- x_i: Received (mixed) signals.
- y_i: Estimated sources.

Goal: $y_i = s_i$

- Is it possible? Isn't it ill-posed?

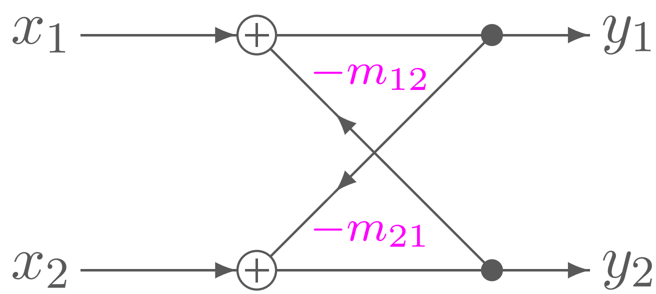
✓ Observation in 1982: The angular position ($p(t)$) and the angular velocity ($v(t) = dp(t)/dt$) of a joint is represented by two nervous signals $f_1(t)$ and $f_2(t)$, each one is a linear combination of position and velocity:

$$\begin{cases} f_1(t) = a_{11}p(t) + a_{12}v(t) \\ f_2(t) = a_{21}p(t) + a_{22}v(t) \end{cases}$$

At each instant the nervous system knows $p(t)$ and $v(t) \Rightarrow$

$p(t)$ and $v(t)$ must be recoverable only from $f_1(t)$ and $f_2(t)$

Part I Herault and Jutten (HJ) Algorithm



- Presented in GRETSI'85, COGNITAVA'85 and Snowbird'86.
- Choosing m_{12} and m_{21} correctly results in separation:

$$\begin{cases} y_1 = x_1 - m_{12} y_2 \\ y_2 = x_2 - m_{21} y_1 \end{cases} \rightarrow \mathbf{y} = \mathbf{x} - \mathbf{M}\mathbf{y} \rightarrow \mathbf{y} = (\mathbf{I} + \mathbf{M})^{-1}\mathbf{x}$$

- Main Idea: $E\{f(y_1)g(y_2)\} = 0 \Leftrightarrow$ Independence (ICA)
- The algorithm:

$$\boxed{\begin{aligned} m_{12} &\leftarrow m_{12} - \mu f(y_1)g(y_2) \\ m_{21} &\leftarrow m_{21} - \mu f(y_2)g(y_1) \end{aligned}}$$

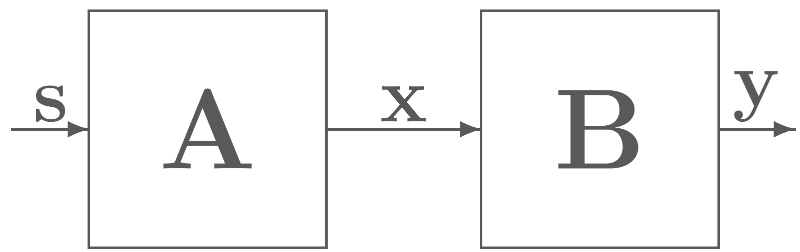
Growing up interest

- mid 80's–mid 90's: Slow development:
  - Neural networks scientists were challenging some other topics!
  - Mainly French scientists.
  - 1988: J.-L Lacoume work based on cumulant.
  - Starting 1989: P. Comon and J.-F. Cardoso's papers.
- 1994: Bell & Sejnowsky work.
- From mid 90's: Exploring interest.
- 1999: First international conference, ICA'99 (France), collecting more than 100 researchers.
- ICA2000 (Finland), ICA2001 (USA), ICA2003 (Japan) and upcoming ICA2004 (Spain).

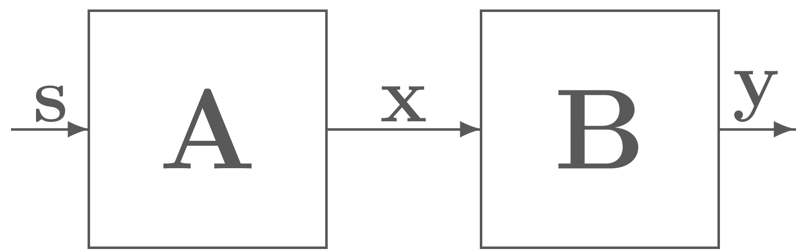
Part II

Separability

Linear (instantaneous) mixtures

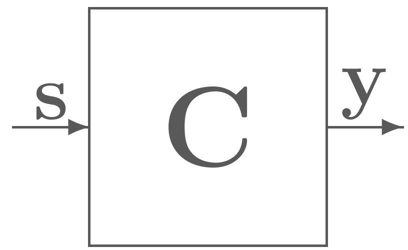


- $\mathbf{s} = (s_1, \ldots, s_N)^T$: source vector.
- $\mathbf{x} = (x_1, \ldots, x_N)^T$: observation vector.
- $\mathbf{y} = (y_1, \ldots, y_N)^T$: output vector.
- $\mathbf{x} = \mathbf{A}\mathbf{s}$: unknown mixing system.
- $\mathbf{y} = \mathbf{B}\mathbf{x}$: separating system.



- Main assumption: The sources (s_i's) are statistically *independent*.
- Separability: Does the independence of the outputs (ICA) imply the separation of the sources (BSS)?

Counter-example



- s_1 and s_2 independent $\sim N(0,1)$.
- C an orthonormal (rotation) matrix.
- $\mathbf{R_y} = E\left\{\mathbf{y}\mathbf{y}^T\right\} = \mathbf{C}\mathbf{R_s}\mathbf{C}^T = \mathbf{C}\mathbf{C}^T = \mathbf{I}$
  $\Rightarrow y_1$ and y_2 are independent.

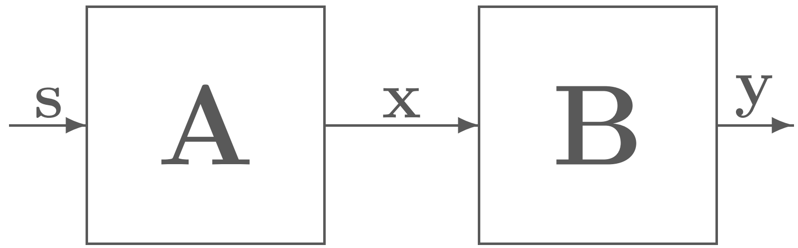
Part II Darmois' Theorem (1947)

$$\begin{cases} y_1 = a_1 s_1 + a_2 s_2 + \cdots + a_N s_N \\ y_2 = b_1 s_1 + b_2 s_2 + \cdots + b_N s_N \end{cases}$$

- s_i's are independent
- y_1 and y_2 are independent
- If for an i we have $a_i b_i \neq 0 \Rightarrow s_i$ is Gaussian.

$$\Downarrow$$

If y_1 and y_2 are independent, a Non-Gaussian source cannot be present in both of them.

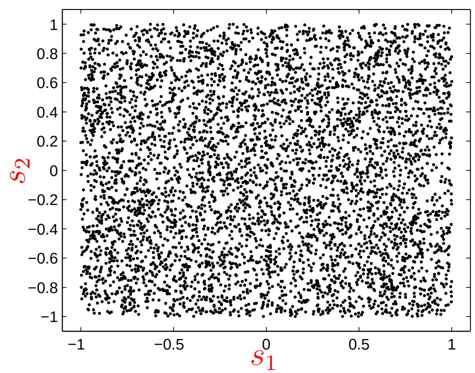


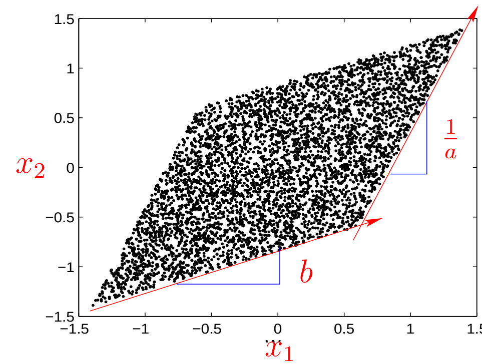
✓ Linear mixtures are *separable*, provided that there is *no more than 1 Gaussian source* ([Comon 91 & 94] inspired from [Darmois 1947]):

Independence of the outputs	$\Longleftrightarrow$	Separation of the sources

✓ Indeterminacies are trivial: Permutation and Scale.

Part II Geometric Interpretation [Puntonet *et. al.* GRETSI 95]





Bounded sources:

$$p_{s_1 s_2}(s_1, s_2) = p_{s_1}(s_1) p_{s_2}(s_2)$$
$$p_{s_2|s_1}(s_2|s_1) = p_{s_2}(s_2)$$

$\Rightarrow$ The distribution of (s_1, s_2) forms a rectangular region

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad \mathbf{A} = \begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix}$$

- Non-Gassianity, isn't it too restrictive?
  - Many practical signals (speech, PSK, bounded signals, . . .) are not Gaussian.
  - Cramer's Theorem:
    - $X = X_1 + X_2 + \cdots + X_N$.
    - X_i's independent.
    - X is Gaussian $\Rightarrow$ All X_i's must be Gaussian.
- Gaussian sources can be separated if there is some time dependence (non-iid) or non-stationarity.

Question: Can we use decorrelation (2nd order independence) for source separation? $\rightarrow$ NO!

Example:

- x_i's are decorrelated ($\mathbf{R_x} = \mathbf{I}$).
- $\mathbf{B}$ any orthogonal (rotation) matrix.
- $\mathbf{y} = \mathbf{Bx} \Rightarrow \mathbf{R_y} = \mathbf{BR_xB}^T = \mathbf{I} \Rightarrow y_i$'s decorrelated.

✓ Decorrelation property remains unchanged under any orthogonal mixing matrix.

Principle Component Analysis (PCA) = Karhunen-Loève Transform = Hotelling transform = Whitening

- $\mathbf{R}_{\mathbf{x}} = E\left\{\mathbf{x}\mathbf{x}^T\right\}$: Covariance matrix of $\mathbf{x}$.
- $\mathbf{E}$ and $\mathbf{\Lambda}$: Eigenvector and eigenvalue matrices of $\mathbf{R}_{\mathbf{x}}$.
- $\mathbf{W} \triangleq \mathbf{E}^T$: The whitening matrix.

$$\mathbf{y} = \mathbf{W}\mathbf{x} \Rightarrow \mathbf{R}_{\mathbf{y}} = \mathbf{\Lambda} \text{ (diagonal)}$$
$$\Rightarrow y_i\text{'s are decorrelated.}$$

- PCA is NOT sufficient for ICA (it leaves an unknown rotation).

Part II Another interpretation

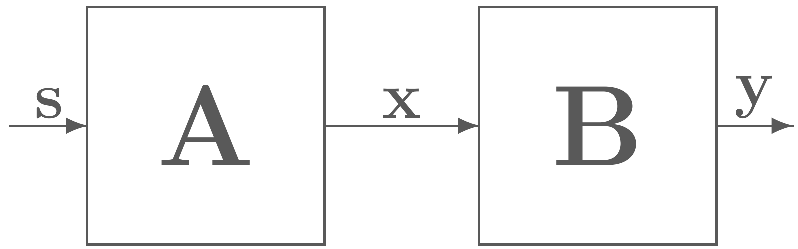
$$\mathbf{y} = \mathbf{B}\mathbf{x} \rightarrow \mathbf{B} = \begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix}$$

✓ 2 unkowns (a and b) must be determined.

✓ Decorrelation property ($E\left\{y_1 y_2\right\} = 0$) gives only 1 equation $\Rightarrow$ Not sufficient.

$\Rightarrow$ Decorrelation (2nd order independence) is not sufficient $\Rightarrow$ Higher Order Statistics (HOS).

Summary of Part II



- Linear mixtures can be separated (At most 1 Gaussian source).
- Remaining indeterminacies: Scale, Permutation.
- Output independence is sufficient for source separation.
- Independence cannot be reduced to decorrelation.

Part III

Some famous approaches for solving BSS problem

- 4th order independence is sufficient.
- Higher order characteristics of a random variable is usually described using its cumulants.
- cumulants: Coefficients of Taylor series of the second characteristic function $\Psi_x(s) = \ln \Phi_x(s) = \ln E\{e^{sx}\}$.
- Cross-cumulants: Coefficients of Taylor series of $\Psi_{x_1 x_2}(s_1, s_2) = \ln E\{e^{s_1 x_1 + s_2 x_2}\}$.
- 4th order independence $\equiv$ Cancelling Cum$_{13}(y_1, y_2)$, Cum$_{22}(y_1, y_2)$ and Cum$_{31}(y_1, y_2)$.
- Requires non-zero 4th order statistics, Only for linear mixtures.

Part III Mutual Information, an independence criterion

Independence of $\mathbf{x} = (x_1, \ldots, x_N)^T \Leftrightarrow p_\mathbf{x}(\mathbf{x}) = \prod_{i=1}^N p_{x_i}(x_i)$

$$I(\mathbf{x}) = \mathrm{KL}\left(p_\mathbf{x}(\mathbf{x}) \| \prod_{i=1}^N p_{x_i}(x_i)\right) = \int_\mathbf{x} p_\mathbf{x}(\mathbf{x}) \ln \frac{p_\mathbf{x}(\mathbf{x})}{\prod_i p_{x_i}(x_i)} d\mathbf{x}$$
$$= \sum_i H(x_i) - H(\mathbf{x})$$

H Shannon's entropy $\rightarrow H(\mathbf{x}) = -E\left\{p_\mathbf{x}(\mathbf{x})\right\}$

Main property:

- $I(\mathbf{x}) \geq 0$.
- $I(\mathbf{x}) = 0$ iff $x_1, \ldots, x_N$ are independent.

Part III Minimizing output mutual information

- $\frac{\partial}{\partial \mathbf{B}} I(\mathbf{y}) = E\left\{\boldsymbol{\psi}_{\mathbf{y}}(\mathbf{y})\mathbf{x}^T\right\} - \mathbf{B}^{-T}$.
  - $\boldsymbol{\psi}_{\mathbf{y}}(\mathbf{y}) \triangleq (\psi_{y_1}(y_1), \ldots, \psi_{y_N}(y_N))^T$.
  - $\psi_{y_i}(y_i) \triangleq -\frac{d}{dy_i} \ln p_{y_i}(y_i)$.
- Steepest descent: $\mathbf{B} \leftarrow \mathbf{B} - \mu \frac{\partial}{\partial \mathbf{B}} I(\mathbf{y})$.
- Equivarient algorithm [Cardoso&Laheld 96]:

$$\mathbf{B} \leftarrow \mathbf{B} - \mu \nabla_{\mathbf{B}} I(\mathbf{y}) \mathbf{B}$$

  - $\nabla_{\mathbf{B}} I(\mathbf{y}) = \frac{\partial}{\partial \mathbf{B}} I(\mathbf{y}) \mathbf{B}^T = E\left\{\boldsymbol{\psi}_{\mathbf{y}}(\mathbf{y})\mathbf{y}^T\right\} - \mathbf{I}$.
- Not applicable for more complicated mixtures ($I(\mathbf{y} + \boldsymbol{\Delta}) - I(\mathbf{y}) = ?$).

Part III Some definitions

- Score function of a random variable x:

$$\psi_x(x) \triangleq -\frac{d}{dx} \ln p_x(x)$$

- For a random vector $\mathbf{x} = (x_1, \ldots, x_N)^T$:
  - Marginal Score Function (MSF):

$$\boldsymbol{\psi}_{\mathbf{x}}(\mathbf{x}) \triangleq (\psi_{x_1}(x_1), \ldots, \psi_N(x_N))^T, \quad \psi_i(x_i) \triangleq -\frac{d}{dx_i} \ln p_{x_i}(x_i)$$

  - Joint Score Function (JSF):

$$\boldsymbol{\varphi}_{\mathbf{x}}(\mathbf{x}) \triangleq (\varphi_1(x_1), \ldots, \varphi_N(x_N))^T, \quad \varphi_i(\mathbf{x}) \triangleq -\frac{\partial}{\partial x_i} \ln p_{\mathbf{x}}(\mathbf{x})$$

  - Score Function Difference (SFD):

$$\boldsymbol{\beta}_{\mathbf{x}}(\mathbf{x}) \triangleq \boldsymbol{\psi}_{\mathbf{x}}(\mathbf{x}) - \boldsymbol{\varphi}_{\mathbf{x}}(\mathbf{x})$$

Part III *Differential* of mutual information

$$I(\mathbf{x}+\boldsymbol{\Delta}) - I(\mathbf{x}) = E\left\{\boldsymbol{\Delta}^T \boldsymbol{\beta}_{\mathbf{x}}(\mathbf{x})\right\} + o(\boldsymbol{\Delta})$$

For a differentiable multi-variate function:

$$f(\mathbf{x}+\boldsymbol{\Delta}) - f(\mathbf{x}) = \boldsymbol{\Delta}^T \cdot (\nabla f(\mathbf{x})) + o(\boldsymbol{\Delta})$$

SFD can be called the stochastic gradient of the mutual information.

✓ Maximizing Non-Gaussianity of the outputs.

- $x_1 = a_{11}s_1 + a_{12}s_2 + \cdots + a_{1N}s_N$: each x_i is 'more Gaussian' than all sources.
- $y_1 = b_{11}x_1 + b_{12}x_2 + \cdots + b_{1N}x_N$: Determine b_{1i}'s to produce as non-Gaussian as possible $y_1 \Rightarrow$ Separation.
- Measure of non-Gaussianity: Neg-entropy.
- Example: FastICA algorithm [Hyvärinen 99].

✓ Second order approaches (applicable for Gaussian sources, too):

- Exploiting time correlation
  - $E\{y_1(n)y_2(n)\} = 0$ and $E\{y_1(n)y_2(n-1)\} = 0$.
  - Requires time correlation (non-applicable for iid sources).
- Exploiting non-stationarity: Joint diagonalization of Covariance matrix [Pham 2001].
  - Requires non-stationarity.

✓ Non-separable if ALL these three properties:

- Gaussian.
- iid.
- stationary.

Summary of Part III

- Algorithms based on output independence:
  - Cancelling 4th order cross-cumulants.
  - Minimizing mutual information.
- Algorithms based on non-Gaussianity.
- Second order algorithms:
  - Algorithms based on time correlation.
  - Algorithms based on non-stationarity.

Part IV

Extensions to ICA

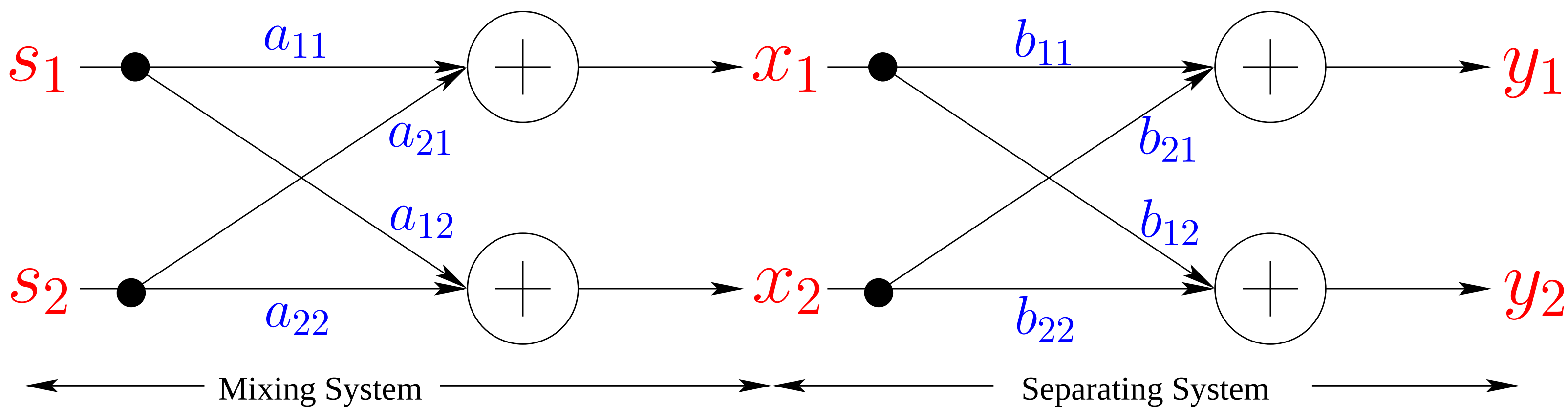
Part IV Extensions to linear instantaneous mixtures

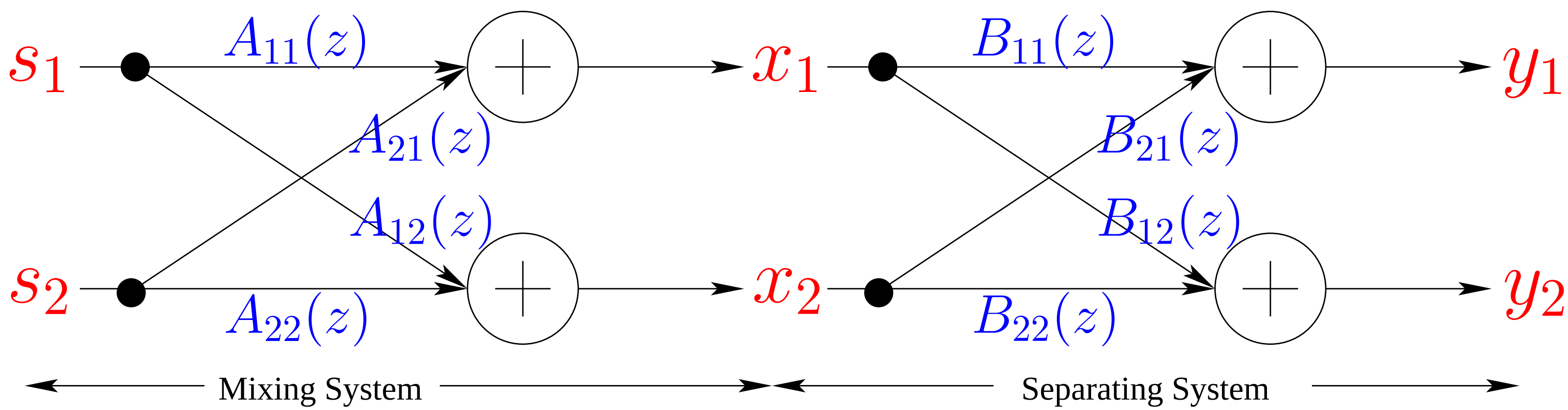
- Complex signals.
- Noisy ICA:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$$

- Different number of sources and sensors:
  - Overdetermined mixtures.
    - Estimating number of sources?
  - Underdetermined mixtures.

Part IV Convolutive Mixtures



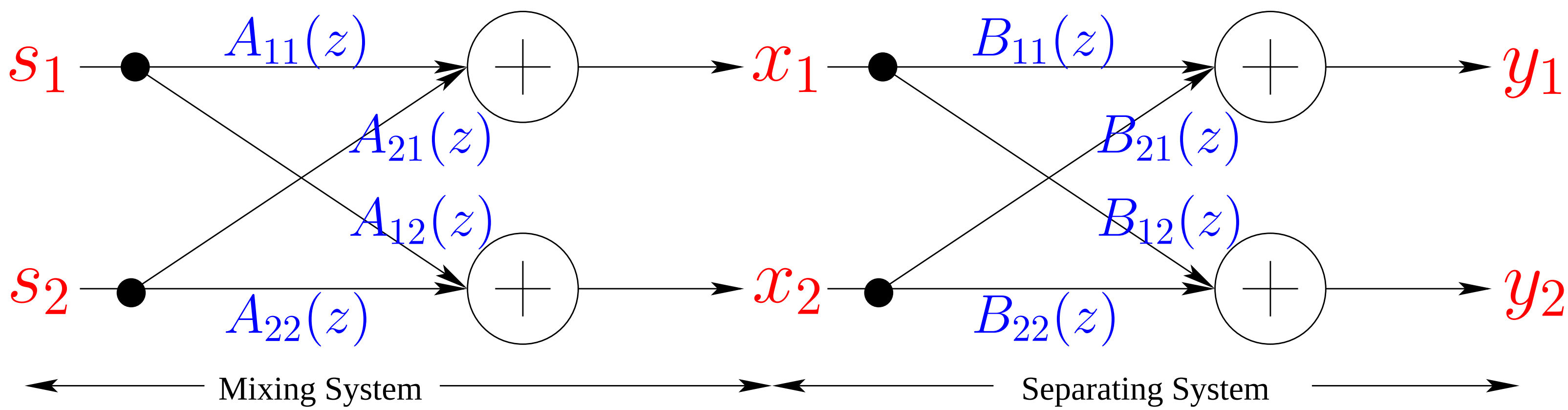


- ✓ Separation system:

$$\mathbf{y}(n) = \mathbf{B}_0\mathbf{x}(n) + \mathbf{B}_1\mathbf{x}(n-1) + \cdots + \mathbf{B}_M\mathbf{x}(n-M)$$

- ✓ Extension to the Widrow's noise canceller system.

Part IV Convolutive Mixtures



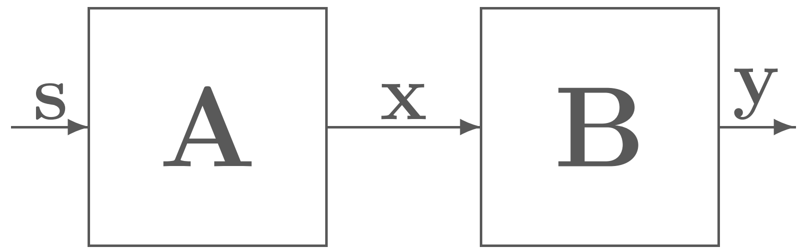
- ✓ Separation system:

$$\mathbf{y}(n) = \mathbf{B}_0\mathbf{x}(n) + \mathbf{B}_1\mathbf{x}(n-1) + \cdots + \mathbf{B}_M\mathbf{x}(n-M)$$

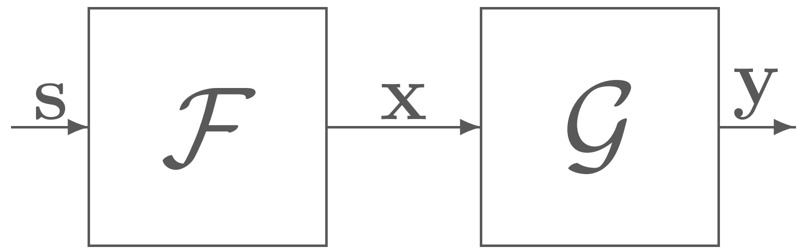
- ✓ Extension to the Widrow's noise canceller system.

- ✓ Convolutive mixtures are separable, too [Yellin, Weinstein, 95]: Output independence $\rightarrow$ Separation.

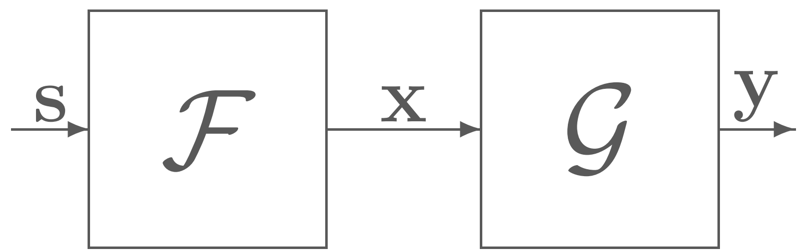
Part IV Non-linear Mixtures



Part IV Non-linear Mixtures



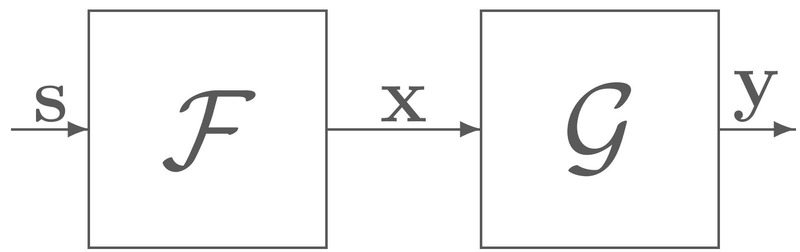
Part IV Non-linear Mixtures



✓ In general, non-linear mixtures are not separable:

Output Independence $\not\Rightarrow$ source separation

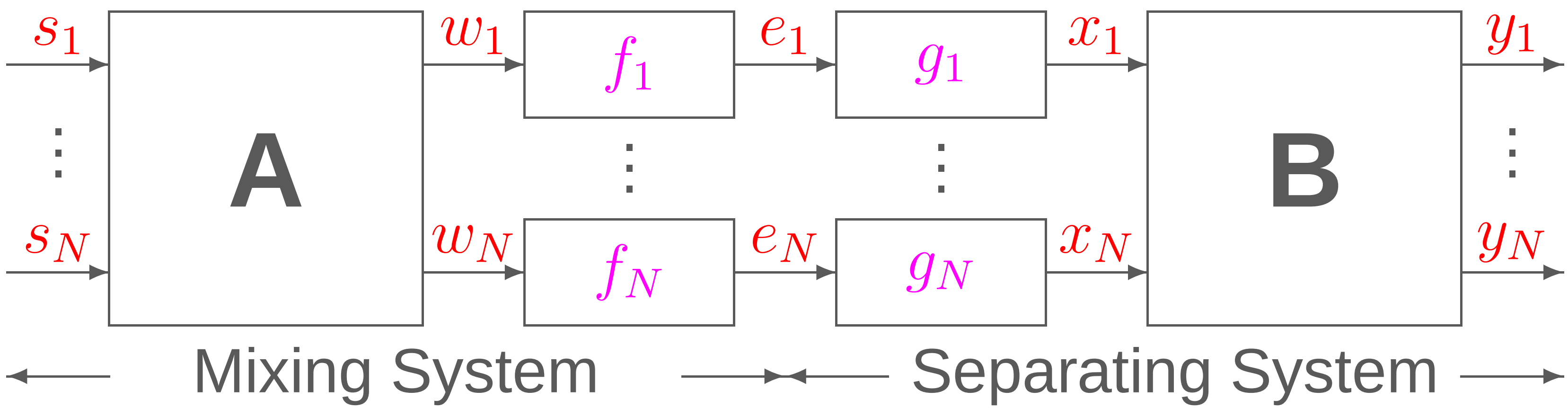
✓ Independence is not strong enough for source separation.



How to overcome this problem?

- Regularization techniques (smoothness)?
- Structural constraints
- Others (temporal correlation? non-stationarity?)

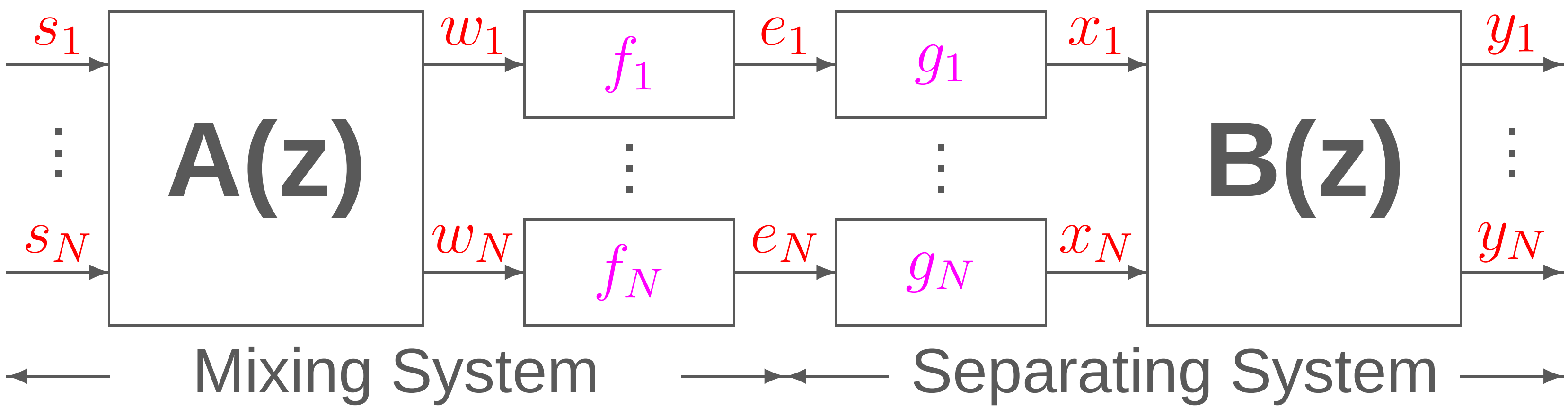
PNL (Post Non-Linear) mixtures



✓ Separability theorem [Taleb & Jutten, IEEE trans. SP, 99]:
The outputs are independent iff:

- $g_i = f_i^{-1}$
- $\mathbf{BA} = \mathbf{PD}$

Provided that: The sources are really mixed (at least 2 non-zero entries in each row of $\mathbf{A}$).



Separability:

$$\mathbf{s} = \begin{bmatrix} \vdots \\ \mathbf{s}^T(n-1) \\ \mathbf{s}^T(n) \\ \mathbf{s}^T(n+1) \\ \vdots \end{bmatrix}, \mathbf{e} = \begin{bmatrix} \vdots \\ \mathbf{e}^T(n-1) \\ \mathbf{e}^T(n) \\ \mathbf{e}^T(n+1) \\ \vdots \end{bmatrix}, \bar{\mathbf{A}} = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \mathbf{A}_{n+1} & \mathbf{A}_n & \mathbf{A}_{n-1} & \cdots \\ \cdots & \mathbf{A}_{n+2} & \mathbf{A}_{n+1} & \mathbf{A}_n & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

$$\mathbf{A}(z) = \sum_n \mathbf{A}_n z^{-n}, \quad \boxed{\mathbf{e} = \mathbf{f}\left(\bar{\mathbf{A}}\mathbf{s}\right)}$$

Part V

Applications, my works and perspectives

Part V Applications

- Feature Extraction.
- Image denoising (using noisy ICA methods).
- Medical engineering applications (ECG, EEG, MEG, Artifact separation).
- Telecommunications (Blind Channel Equalization, CDMA).
- Financial applications.
- Audio separation.
- Seismic applications.
- Astronomy.

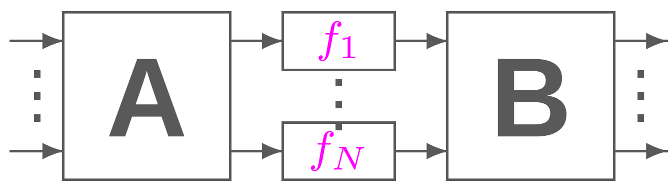
Part V My works, mainly at my PhD thesis

- CPNL mixtures.
- Gradient of mutual information (SFD):
  - General approach for any (separable) parametric model.
  - Gradient approach.
  - Minimization-Projection approach.
  - Special cases: Linear, convolutive and PNL.
- Proof of separability of PNL mixtures.
- A geometric method for separating PNL mixtures (compensating sensors' nonlinearities before separation).
- Post Convolutive mixtures and their properties.
- Even smooth non-linear systems may preserve the independence.
- (Not at my PhD thesis) Blind estimation of a Wiener telecommunication channel (linear channel + nonlinear receiver).
- Manuscript downloadable from:

  `http://www.lis.inpg.fr/theses`

Part V Perspectives

- Continuation of my previous works
  - Writing 2 papers.
  - Adaptive algorithms.
  - Underdetermined mixtures: it seems that the minimization-projection approach can be used for identifying (but not separating) such systems.
  - Working on PNL-L mixtures:



  - PNL mixtures: Compensating sensor non-linearities before separation.
  - Further work on developed algorithms (improvements, convergence analysis, . . .)
  - Searching for £nding better (maybe optimal) SFD estimators.
  - A few other small ideas.
- Working on audio source separation.