6. Nonlinear mixtures

Identifiability

Three ideas of regularization

PNL mixtures and algorithms

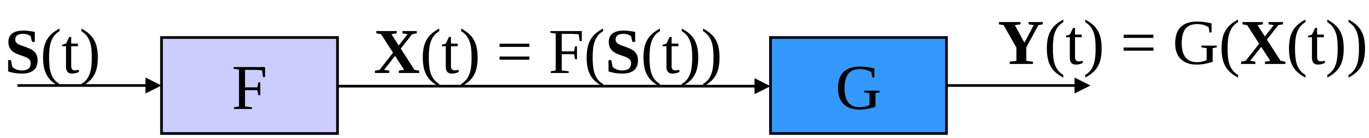
1. Introduction
Model and problem

- Multidimensional (P) observations (sensor array, satellite antenna, microphone array, etc.) are mixtures of N *independent* sources $x_i(t) = f_i(s_1(t), \cdots, s_N(t)) + n_i(t), \quad i = 1, \cdots, P$
- Denoting $\mathbf{x}(t) = \left[x_1(t), \cdots, x_P(t)\right]^T$ and $\mathbf{s}(t) = \left[s_1(t), \cdots, s_N(t)\right]^T$:

$$\mathbf{x}(t) = \mathbf{F}(\mathbf{s}(t)) + \mathbf{n}(t)$$

  where **n**(t) is a noise, independent of **s**(t), **F**(.) and **s**(t) are unknown.
- Source separation consists in extracting the sources $s_j(t)$ from the observations $x_i(t)$ by identifying an inverse model G:

$$\mathbf{y}(t) = \mathbf{G}\left[\mathbf{F}(\mathbf{s}(t) + \mathbf{n}(t)\right] = \mathbf{H}(\mathbf{s}(t) + \mathbf{n}(t)) \approx \mathbf{s}(t)$$



I. Introduction
Separability

- The key question is *separability*.

- In other words: Is independence sufficient for leading to source separation (i.e. for estimating G) ?

$$ICA \Leftrightarrow BSS\ ?$$

1. Darmois' results
Nonlinear mappings

- **Nonlinear invertible mapping *F* are not separable**
  - Let s_i, s_j be 2 independent variables, the random variables $f_i(s_i), f_j(s_j)$ are independent too, provided that f_i, f_j are invertible mappings

i.e. sources can be separated, but only up to an invertible NL mapping

  - Moreover, there exists many Mixing Mappings which Preserves Independence (**MMPI**), i.e. for these mappings, ICA does not imply BSS.
  - See (Darmois, 1953) for general construction, and the following example.

1. Source separation principles Separability (2/6)

- Separability: general case, i.e. ***F*** is a nonlinear invertible mapping
- NL Indeterminacies
  - if $S_i = k_i(S_i^*)$, the random variables S_i^* are independent too, and ***E = F(S) = F(K(S*))***.

    I.e., sources are separable up to an invertible NL mapping.
  - Consider the mapping Y=G(X), and assume densities exist:

    $$p_Y(\mathbf{y}) = p_X(\mathbf{x})\left|\det J_G\right|^{-1}(\mathbf{x})$$

    Without loss of generality, we also assume Y is uniform in $[0,1]^n$:

    We are looking for a mapping G such that Y is independent, i.e.:

    $$\left|\det J_G\right|(\mathbf{x}) = p_X(\mathbf{x})$$

1. Source separation principles
Separability (3/6)

- For example, we can look for solutions g_i monotonous in u_i satisfying:

$$\begin{cases} g_1(\mathbf{u}) = g_1(u_1) \\ g_2(\mathbf{u}) = g_2(u_1, u_2) \\ \vdots \\ g_n(\mathbf{u}) = g_n(u_1, u_2, \cdots, u_n) \end{cases}$$

It then leads to :

$$\begin{vmatrix} \dfrac{\partial g_1}{\partial u_1} & 0 & \cdots & 0 \\ \dfrac{\partial g_2}{\partial u_1} & \dfrac{\partial g_2}{\partial u_2} & 0 & \cdots \\ \vdots & \cdots & & \vdots \\ \dfrac{\partial g_n}{\partial u_1} & \dfrac{\partial g_n}{\partial u_2} & \cdots & \dfrac{\partial g_n}{\partial u_n} \end{vmatrix} = \prod_i \frac{\partial g_i}{\partial u_i}$$

1. Source separation principles Separability (4/6)

- The last condition implies:

$$\prod_i \frac{\partial g_i}{\partial x_i} = p_X(\mathbf{x}) = p(x_1)p(x_2/x_1)\cdots p(x_n/x_1,\cdots,x_{n-1})$$

- A solution is then:

$$\begin{cases} g_1 = F_{X_1} \\ g_2 = F_{X_2/X_1} \\ \vdots \\ g_n = F_{X_n/X_1,\cdots,X_{n-1}} \end{cases}$$

- This idea for constructing G is originally due to Darmois, 1951. A quite similar construction is used by Hyvärinen and Pajunen (Neural Networks, 1999).

1. Source separation principles
Separability (5/6)

- 2-D example
  - X and Y are two independent Gaussian variables with joint pdf

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

  - Consider the mapping, and its Jacobian matrix:

$$\begin{cases} X = r\cos\theta \\ Y = r\sin\theta \end{cases} \qquad J = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

  - The joint pdf of the new variables is:

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

1. Source separation principles
Separability (6/6)

- Statistical independence is not sufficient for insuring separation in non linear mixtures
  - there exists an infinity of invertible mappings which preserves independence without having a diagonal Jacobian matrix
  - if the mapping has diagonal Jacobian matrix, source can be only recovered up to a NL mapping

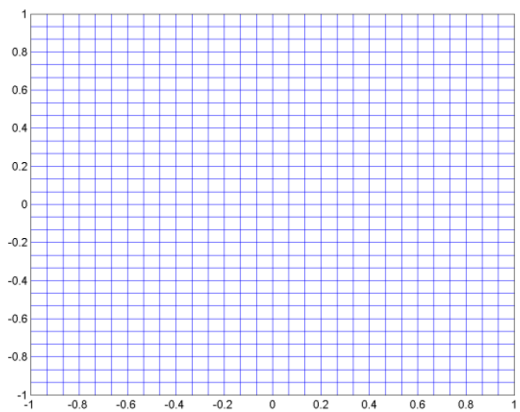
- Separating NL mixtures requires structured constraints or regularization techniques

2. How regularizing for avoiding MMPI ?

- Many nontrivial mapping preserving independence: one has to reduce the indeterminacy by “regularizing”
- We consider three ways:
  - Restrict G to smooth mappings (Almeida et al., Wu et Zurada), or des Bayesian approach (Valpola et al.),
  - Restrict G by structural constraints (Taleb & Jutten, PNL mixtures, 1999 ; Kagan et al., Mappings satisfying addition theorem, 1973 and Eriksson & Koivunen, Eusipco’02, Toulouse)
  - Exploiting particular source properties : bounded sources (Babaie-Zadeh & Jutten, Eusipco’02, Toulouse), Markov sources (Hosseini & Jutten, *IEEE NPL*, 2003)
- Other ways ?

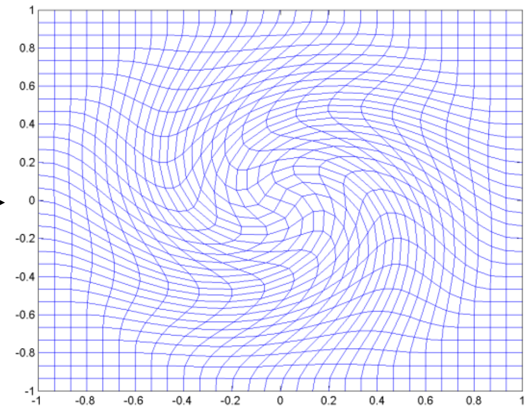
2.1. Smooth mappings ?

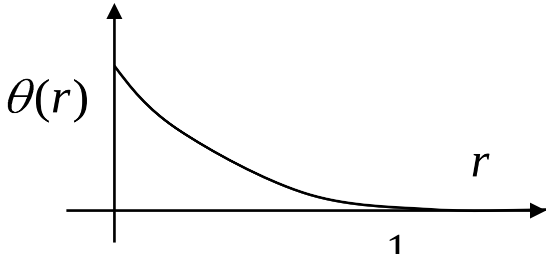
- A smooth mapping can preserve independence without separating the sources.
- As an example (M. Babaie-Zadeh, PhD thesis ; Jutten, Babaie-Zadeh, Hosseini, Signal Processing, 2003)



$$\xrightarrow{\begin{pmatrix} \cos(\theta(r)) & -\sin(\theta(r)) \\ \sin(\theta(r)) & \cos(\theta(r)) \end{pmatrix}}$$

$$\theta(r) = \begin{cases} \theta_0(1-r)^n & 0 \le r \le 1 \\ 0 & r \ge 1 \end{cases}$$





2.2. Structural constraints

(Jutten and Taleb, ICA'2000 ; Taleb, IEEE SP, 2002)

- **Definition**: Obvious mappings

  H is an obvious mapping if ***any*** random vector with independent components is transformed by *H* in another random vector with independent components *I*

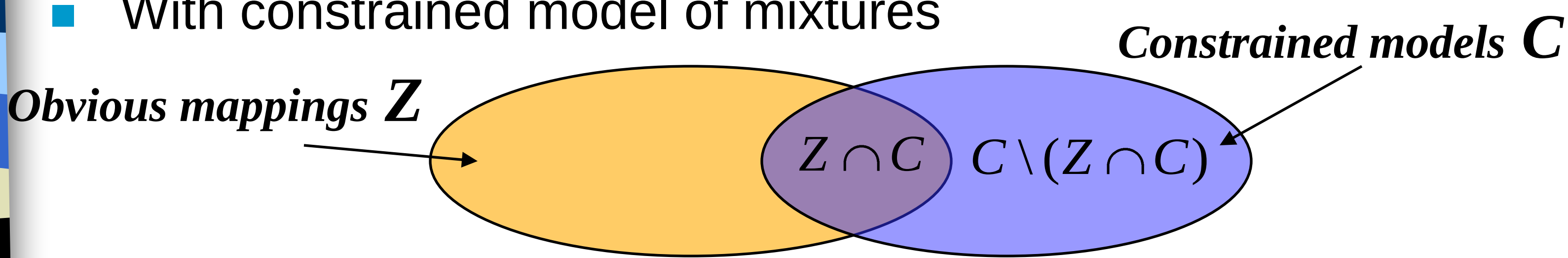
  An obvious mapping is then an independence preserving mapping
- It can be shown that obvious mappings satisfy:

$$H_i(s_1,\cdots,s_n) = h_i(s_{\sigma(i)}), i = 1,\cdots,n$$

- The Jacobian matrix of an obvious mapping is the product of a diagonal matrix and a permutation matrix
- The set of obvious mappings is denoted ***Z***

2.2. Structural constraints

- Darmois's results claim that there is an infinity of non obvious mappings which preserve independence
- With constrained model of mixtures



– If the mapping *H* = *G* o *F* is constrained to belong to a model set ***C*** , the indeterminacies can be reduced, and hopefully cancelled

– Then, consider the set of signal distributions

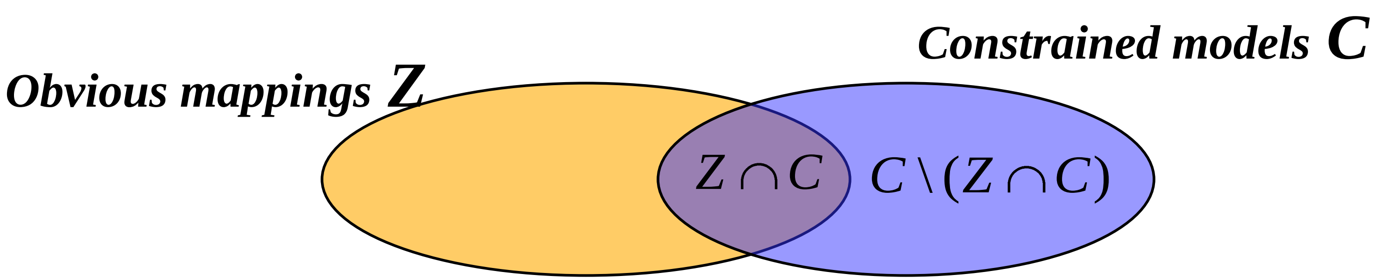
$$\Omega = \{(F_{S_1}, \cdots, F_{S_n}), \text{such that } \exists H \in C \setminus (Z \cap C), H(s) \text{ independent}\}$$

Ω then contains all the distributions which cannot be separated by mapping belonging to ***C***

- Separation is possible for source distributions which do not belong to Ω with indeterminacy $H \in Z \cap C$

2.2. Structural constraints

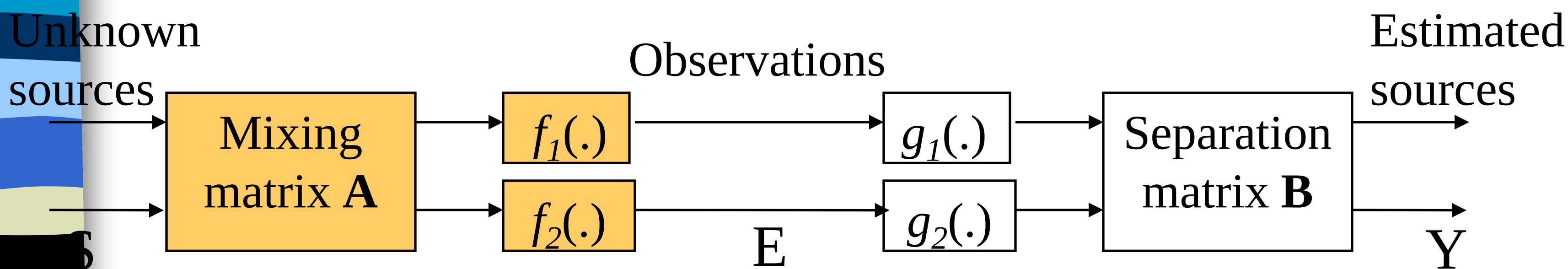
- **Case of regular linear mixtures**
  - C is the set of square matrices
  - $Z \cap C$ is the set of matrices which are the product of a diagonal matrix and a permutation matrix
  - The set Ω is the set of distributions which contains at least 2 Gaussian sources (consequence of the Darmois-Skitovitch theorem)
  - Source separation is then possible provided than there is at most one Gaussian source (avoiding Ω); sources can be restored apart from one permutation and one scale factor.



2.2. Structural constraints
PNL mixtures

- PNL are particular NL mixtures, i.e. with structural constraints

(A. Taleb, C. Jutten, IEEE Trans. SP, Sept. 1999)



- PNL are realistic enough: linear channel + NL sensors and amplifiers

- PNL are separable NL mixtures,
  - If at most, one source is Gaussian, if the mixing matrix has at least 2 non zero entries per row and per column, and if the functions f_i are invertible, then outputs are independent iff $g_i o f_i$ is linear and BA=DP
  - PNL are particular mixtures satisfying addition theorem (see next slides)

2.2. Constrained NL mappings (1/3)

- D-S theorem has been extended to NL function satisfying an addition theorem (Kagan, Linnik, Rao, Communications in Statistics, 1(5),471-474, 1973)

$$f(x+y) = F[f(x), f(y)] = f(x) \circ f(y)$$

  where F is continuous ate least separately on each variable
- Example:

$$f(.) = \tan(.) \qquad \tan(x+y) = \frac{\tan(x)+\tan(y)}{1+\tan(x)\tan(y)}$$

$$F(u,v) = (u+v)/(1+uv)$$

- If (E,o) is an Abelian group, one can define a « multiplication » (*)

$$f(cx) = c * f(x) \text{ or } cf^{-1}(x) = f^{-1}(c*x)$$

$$c_1 f^{-1}(u_1) + \cdots + c_k f^{-1}(u_k) = f^{-1}(c_1 * u_1 \circ c_2 * u_2 \circ \cdots \circ c_k * u_k)$$

- Then, one has the relationships

2.2. Constrained NL mappings (2/3)

- Theorem. Let $X_1,\ldots X_n$ be independent random variables such that

$$E_1 = a_1 * X_1 \circ \cdots \circ a_n * X_n$$
$$E_2 = b_1 * X_1 \circ \cdots \circ b_n * X_n$$

are independent, and where * and o satisfy the above conditions, then $f^{-1}(x_i)$ is normally distributed if $a_i b_i \neq 0$

- The proof is based on application of the D-S theorem to:

$$f^{-1}(E_1) = f^{-1}(a_1 * X_1 \circ \cdots \circ a_n * X_n) = a_1 f^{-1}(X_1) + \cdots + a_n f^{-1}(X_n)$$
$$f^{-1}(E_2) = f^{-1}(b_1 * X_1 \circ \cdots \circ b_n * X_n) = b_1 f^{-1}(X_1) + \cdots + b_n f^{-1}(X_n)$$

- Kagan et al. give a few examples of solutions of the functional equations quoted by Aczel (Lectures in Functional Equations and Their Applications, Academic Press, 1966)

2.2. Constrained NL mappings (3/3)

- PNL mixtures are particular NL mappings satisfying the addition theorem

$$F(u_1,u_2)=\begin{cases} f_1\left(a_{11}f_1^{-1}(u_1)+a_{12}f_2^{-1}(u_2)\right) \\ f_2\left(a_{21}f_1^{-1}(u_1)+a_{22}f_2^{-1}(u_2)\right) \end{cases}$$

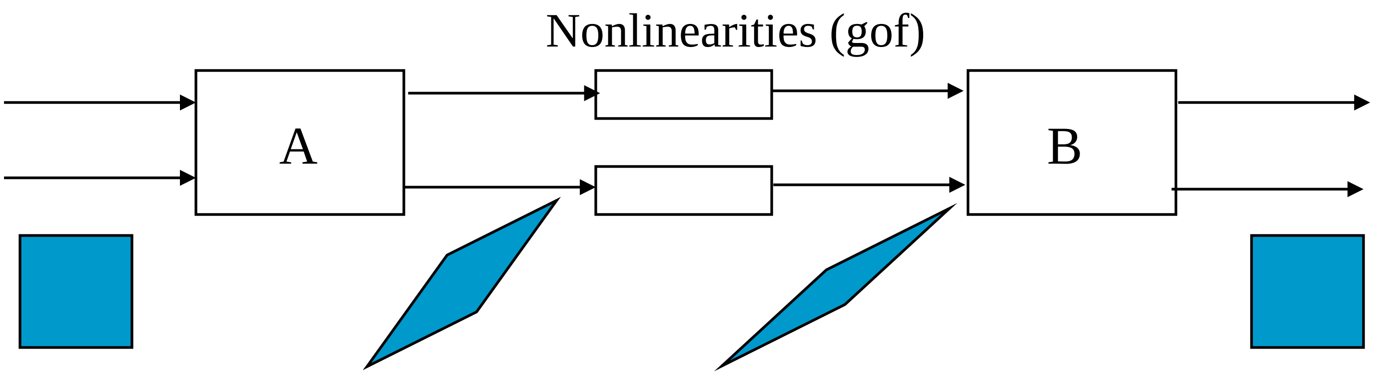
- In fact, denoting

$$s_1 = f_1^{-1}(u_1)\,\text{and}\, s_2 = f_2^{-1}(u_2)$$

$$F(f_1(s_1), f_2(s_2)) = \begin{cases} f_1\left(a_{11}s_1 + a_{12}s_2\right) \\ f_2\left(a_{21}s_1 + a_{22}s_2\right) \end{cases}$$

2.3. Constraints on sources
Bounded sources

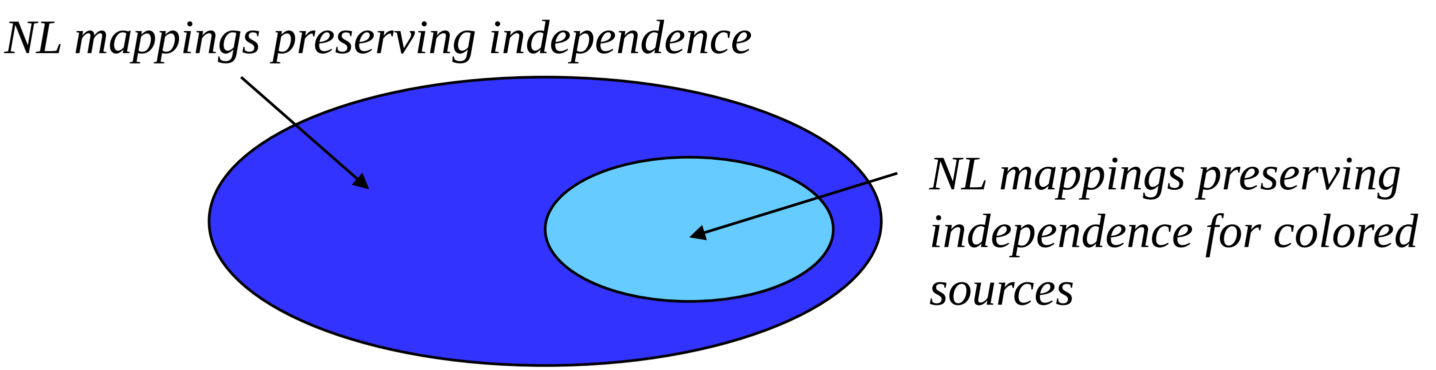
- PNL mixtures of bounded sources (Babaie-Zadeh & Jutten, Eusipco 2002 ; Babaie-Zadeh, PhD Thesis INPG - 2002)



- **Theorem.** A component-wise mapping preserves boundary linearity iff it is a linear mapping
  – New proof of separability for PNL mixtures
  – Algorithm with independent estimations of linear and nonlinear parts

2.3. Constraints on sources
Temporally correlated sources (1/3)

- Hosseini and Jutten (IEEE SP Letters, 2003)
- Sources temporally correlated (e.g. modelled by AR or Markov models)
- The set of NL mixing mappings preserving independence is reduced if sources are non iid, due to stronger independence criterion (independence of random processes)



2.3. Constraints on sources
Temporally correlated sources (2/3)

- One considers the Darmois's construction for mixing mappings preserving independence (for 2 mixtures of 2 sources)

$$\begin{cases} y_1(t) = g_1(x_1(t), x_2(t)) = F_{X_1}(x_1(t)) \\ y_2(t) = g_2(x_1(t), x_2(t)) = F_{X_2/X_1}(x_1(t), x_2(t)) \end{cases}$$

- As a result, $p_{Y_1Y_2}(y_1(t), y_2(t)) = p_{Y_1}(y_1(t))\, p_{Y_2}(y_2(t))$
- Since sources are temporally correlated, one can prove that:

$$p_{Y_1Y_2}(y_1(t+1), y_2(t)) \neq p_{Y_1}(y_1(t+1))\, p_{Y_2}(y_2(t))$$

$$p(F_{X_1}(x_1(t+1), F_{X_2/X_1}(x_1(t)x_2(t)) \neq p(F_{X_1}(x_1(t+1))\, p(F_{X_2/X_1}(x_1(t)x_2(t))$$

2.3. Constraints on sources Temporally correlated sources (3/3)

After some computations, one can show that the relation becomes:

$$\int p_{A/B,C}(a,b,c)\, p_B(b)\, db \neq p_A(a)$$

- Clearly, the left-side term is a function of variables *a* and *c*, and consequently, it is usually different of the right-side term.

Then, the above mapping is no longer MMPI for colored sources;

in other words, coloration allows to restrict the indeterminacy.

Conclusion

- For NL mixtures, independence does not insure source separation
- Regularization is required for reducing solution to trivial mappings
  - Smooth mapping is not a sufficient constraint, although...
  - Structural constraints leads to particular NL separable mixtures
    - Mappings satisfying Addition theorem, PNL mixtures
  - Constraints on the sources can reduce the set of solutions
  - Bayesian approaches, *e.g*. ensemble learning (Valpolla et al.)

- Extension to Nonlinear ICA ?
- Are NL mixtures practically useful ? or are they only a theoretical curiosity ?

3. Source separation principles Independence criterion (1/2)

- It consists in estimating an inverse mapping G which provides independent signals

$$\mathbf{S}(\mathrm{t}) \longrightarrow \boxed{\mathrm{F}} \longrightarrow \mathbf{X}(\mathrm{t}) = \mathrm{F}(\mathbf{S}(\mathrm{t})) \longrightarrow \boxed{\mathrm{G}} \longrightarrow \mathbf{Y}(\mathrm{t}) = \mathrm{G}(\mathbf{X}(\mathrm{t}))$$

- Source separation is based on independence criterion on outputs

$$p_Y(y_1,\cdots,y_n) = \prod_i p_{Y_i}(y_i)$$

- The mutual information is a convenient independence criterion

$$I(Y) = \int p_Y(y_1,\cdots,y_n) \ln\left(\frac{p_Y(y_1,\cdots,y_n)}{\prod_i p_{Y_i}(y_i)}\right) \mathbf{dy} = E\left[\ln\left(\frac{p_Y(y_1,\cdots,y_n)}{\prod_i p_{Y_i}(y_i)}\right)\right]$$

1. Source separation principles Independence criterion (2/2)

- The mutual information writes:

$$I(\mathbf{Y}) = \sum_i H(Y_i) - H(\mathbf{Y}) = \sum_i H(Y_i) - H(\mathbf{X}) + E \ln\left|\det J_G\right|$$

- Estimation equation is then: $\frac{\partial I(Y)}{\partial \Theta} = 0$

  where Θ represents the parameter vector

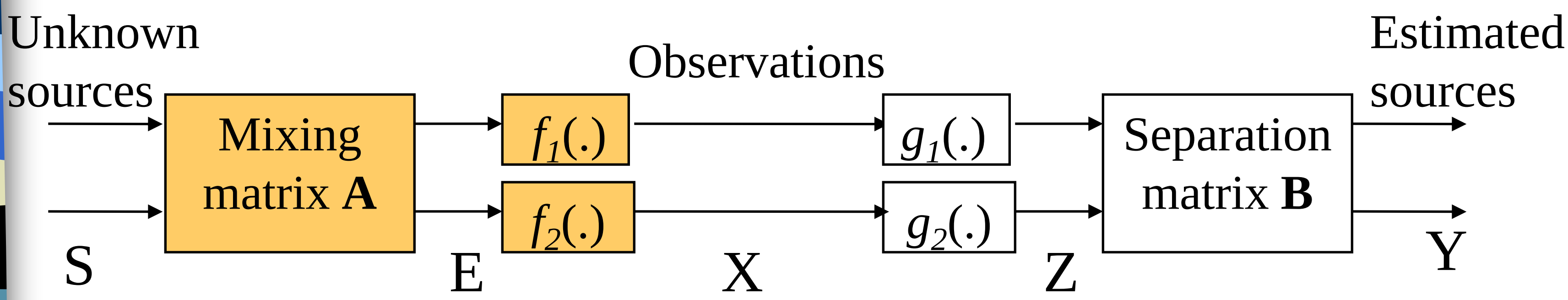
- It leads to the following equations:

$$\frac{\partial I(Y)}{\partial \Theta} = \sum_i \frac{\partial H(Y_i)}{\partial \Theta} + \frac{\partial E \ln\left|\det J_G\right|}{\partial \Theta}$$

$$= -\sum_i E \frac{d \ln p_{Y_i}(y_i)}{dy_i} \frac{\partial y_i}{\partial \Theta} + \frac{\partial E \ln\left|\det J_G\right|}{\partial \Theta} = 0$$

2. Source separation in PNL mixtures Model and Separability

- PNL are particular NL mixtures, i.e. with structural constraints



- PNL are realistic enough: linear channel + NL sensors and amplifiers

- PNL are separable NL mixtures,
  - If at most, one source is Gaussian, if the mixing matrix has at least 2 non zero entries per row and per column, and if the functions f_i are invertible, then outputs are independent iff $g_i o f_i$ is linear and BA=DP

Source separation in PNL mixtures Criterion

- The mutual information writes:

$$
\begin{aligned}
I(\mathbf{Y}) &= \sum_i H(Y_i) - H(\mathbf{Y}) \\
&= \sum_i H(Y_i) - H(\mathbf{X}) - E\ln\left|\det J_G\right| - E\ln\left|\det J_B\right| \\
&= \sum_i H(Y_i) - H(\mathbf{X}) - E\ln\left|\prod_i \frac{\partial g_i}{\partial x_i}(x_i,\theta_i)\right| - \ln\left|\det \mathbf{B}\right|
\end{aligned}
$$

Source separation in PNL mixtures Estimation equation (1/3)

- Deriving I(Y) with respect to B

$$\sum_i E\left[\frac{d \ln p_{Y_i}(y_i)}{dy_i}\frac{dy_i}{d\mathbf{B}}\right] - \mathbf{B}^{-T} = 0$$

$$E\left[\psi_Y(\mathbf{y})\mathbf{y}^T\right] - I = 0$$

- Deriving with respect to θ (parameters of g_i)

$$\sum_i E\left[\frac{d \ln p_{Y_i}(y_i)}{dy_i}\frac{dy_i}{d\theta_k}\right] - \frac{\partial}{\partial \theta_k} E \ln\left|g'_k(x_k, \theta_k)\right| = 0$$

$$E\left[\sum_i \psi_{Y_i}(y_i) b_{ik}\frac{dg_k}{d\theta_k}\right] - E\frac{\partial \ln\left|g'_k(x_k, \theta_k)\right|}{\partial \theta_k} = 0$$

Source separation in PNL mixtures
Estimation equation (2/3)

- Linear part estimation does not require accurate score functions:

$$E\left[\psi_Y(\mathbf{y})\mathbf{y}^T\right] = I \Leftrightarrow \begin{cases} E\psi_{Y_i}(y_i)y_i = 1 \\ E\psi_{Y_i}(y_i)y_j = 0 \quad i \neq j \end{cases}$$

Consequently, if Y_j are zero-mean independent random variables

$$E\psi_{Y_i}(y_i)y_j = E\psi_{Y_i}(y_i)Ey_j = 0 \quad i \neq j$$

- On the contrary, nonlinear part requires an accurate score function estimation :

$$E\left[\sum_i b_{ij}\psi_{Y_i}(Y_i) \middle| Z_j\right] = \psi_{Z_j}(Z_j),\ j = 1,\cdots n$$

with $Z_j = g_j(X_i)$ and $\mathbf{Y} = \mathbf{BZ}$

Source separation in PNL mixtures Estimation equation (3/3)

- Estimation equation requires pdf's or score functions estimates
- pdf's can be estimated using various methods
  - expansion near Gaussianity: Gram-Charlier (Lacoume 91, Comon SP 94, Taleb & Jutten, ESANN 97, Yang et al. SP 98), Edgeworth
  - kernel estimators (Pham IEEE Trans. SP 96, Taleb, Jutten IEEE SP 99)

  … then, score functions are estimated by derivation
- Score function can be estimated directly by minimizing MSE cost (Pham et al. EUSIPCO 92 ; Taleb, Jutten ICANN 97, IEEE SP 99)

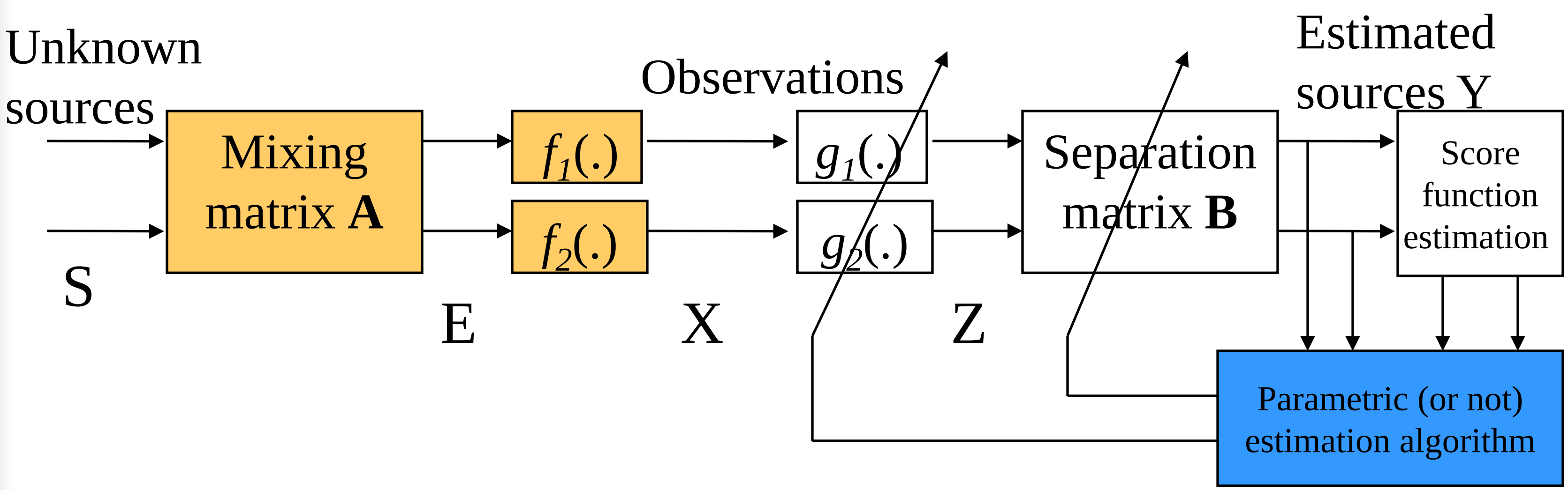
$$J(\mathbf{w}) = \frac{1}{2} E\left[\left(\hat{\psi}_Y(\mathbf{w}, y) - \psi_Y(y)\right)^2\right]$$

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = E\left[\frac{\partial \hat{\psi}_Y(\mathbf{w}, y)}{\partial \mathbf{w}} \left(\hat{\psi}_Y(\mathbf{w}, y) - \psi_Y(y)\right)\right]$$

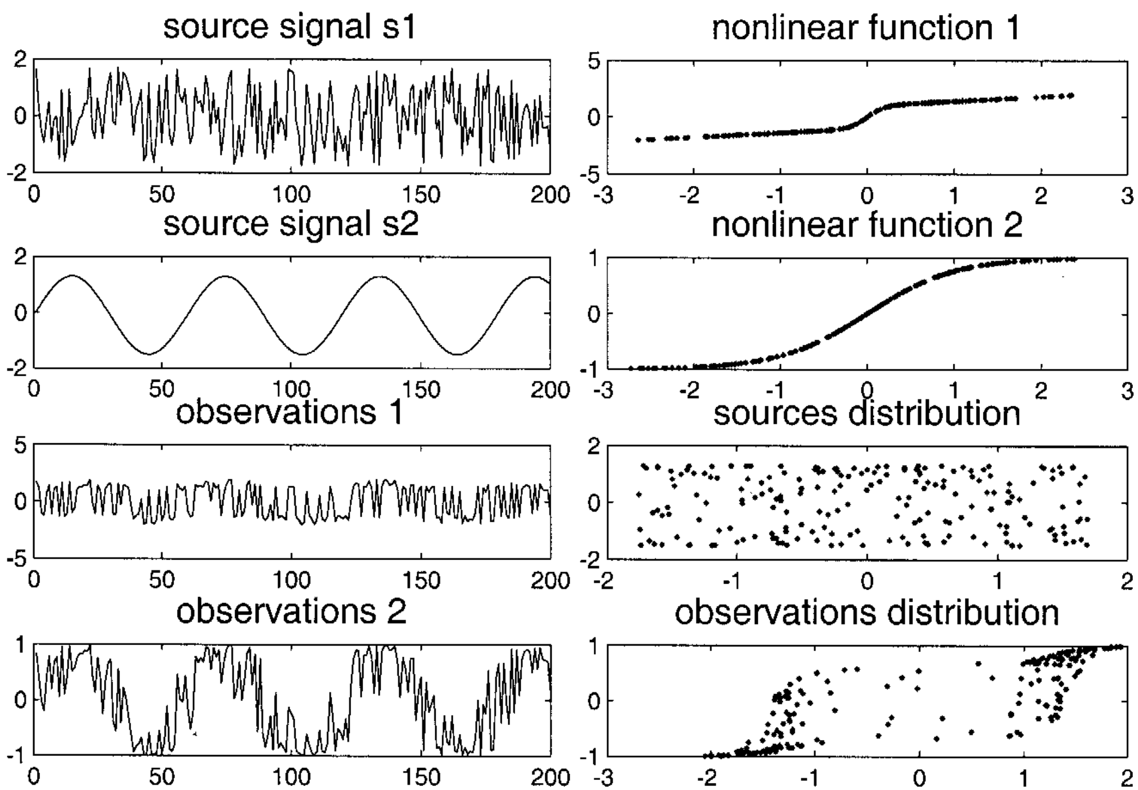
$$= E\left[\hat{\psi}_Y(\mathbf{w}, y) \frac{\partial \hat{\psi}_Y(\mathbf{w}, y)}{\partial \mathbf{w}} + \frac{\partial^2 \hat{\psi}_Y(\mathbf{w}, y)}{\partial y \partial \mathbf{w}}\right]$$

Source separation in PNL mixtures Algorithm

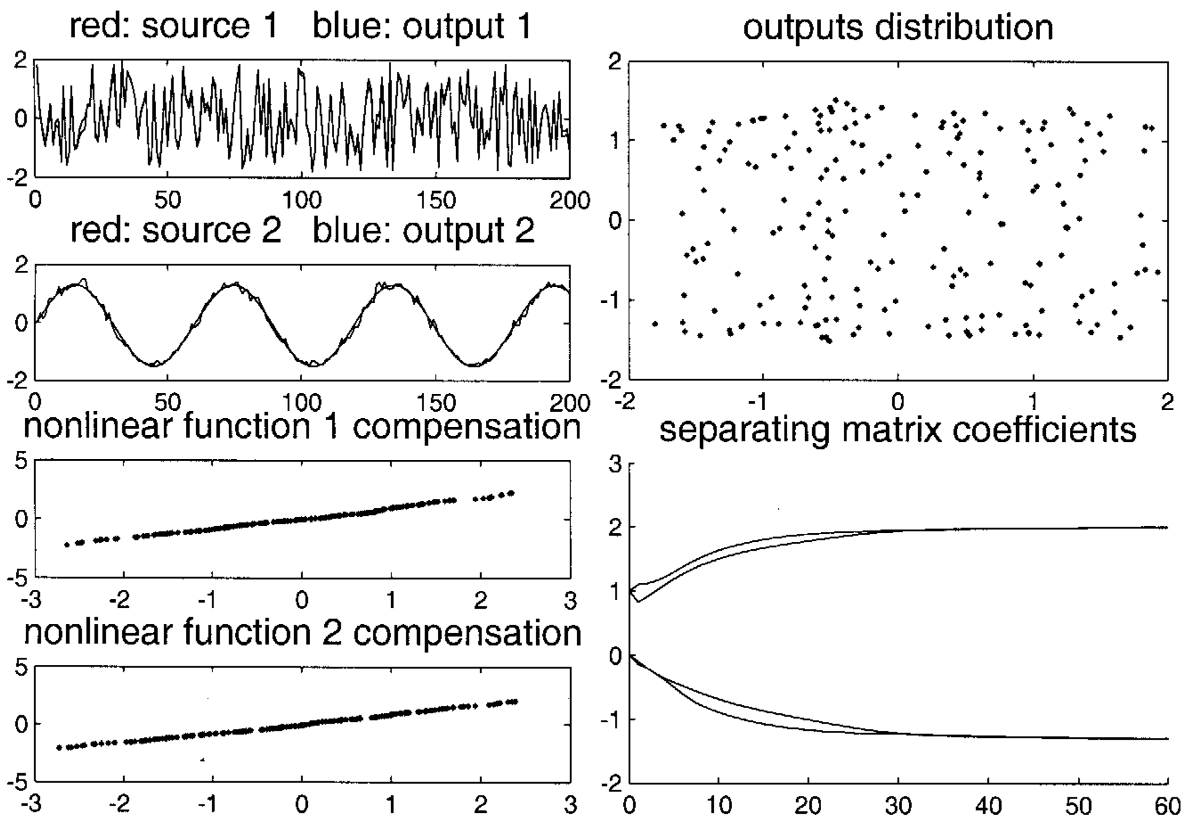
- The algorithm is based on the estimation of 3 parts:
  - marginal score functions of estimated sources,
  - estimation of the nonlinear functions g_i,
  - estimation of the separationg matrix **B**,



Source separation in PNL mixtures Examples (1/2)

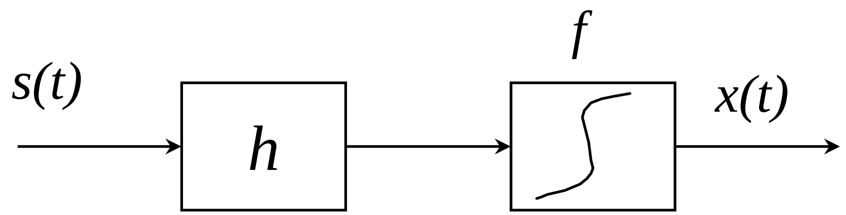


Source separation in PNL mixtures Examples (2/2)

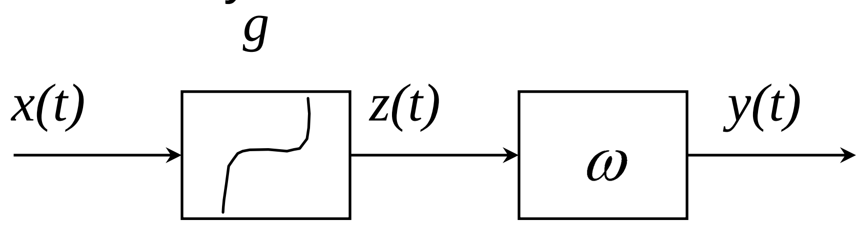


Blind inversion of Wiener systems
The model

- Wiener system



- Hammerstein system



Blind inversion of Wiener systems Classical approaches

- Wiener system is a usual NL model in biology, in satelite communications, etc.

- Classical identification methods for nonlinear systems are based on higher-order cross-correlations
- Usually, input signal is assumed to be iid Gaussian
- If the distortion input is available, the compensation of the nonlinearities is almost straighforward, after identification of the NL

- However, in a real world situation, we don't know either the nonlinear system input or the input distribution

Blind inversion of Wiener systems Wiener/PNL (1/2)

- With the following parameterization:

$$\mathbf{S}(t) = \left[\cdots, s(t-k+1), s(t-k), s(t-k-1), \cdots\right]^T$$

$$\mathbf{X}(t) = \left[\cdots, x(t-k+1), x(t-k), x(t-k-1), \cdots\right]^T$$

$$\mathbf{H} = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & h(k-1) & h(k) & h(k+1) & \cdots \\ \cdots & h(k-2) & h(k-1) & h(k) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

since the scalar input *s*(t) is iid, **S**(t) has independent components and consequently **X**(t) is a mixture of independent sources, the Wiener system is nothing but an infinite dimensionnal postnonlinear mixture: $\mathbf{X}(t) = \mathbf{f}\left(\mathbf{H}\mathbf{S}(t)\right)$

Blind inversion of Wiener systems Wiener/PNL (2/2)

- If the Wiener system satisfies:
  - Subsystems *h* and *f* are unknown and invertible ; *h* can be a nonminimum phase filter
  - The input *s(t)* is an unknown (*a priori*) non Gaussian iid process

  it is equivalent to PNL mixtures, with a particular Toeplitz mixture matrix **H**, with same NL function *f* on each channel

- PNL separability implies Wiener systems inversibility
- PNL separability is only proved for finite dimensions. It is conjectured for infinite dimensions
- Practically, the filter h(k) and its inverse w(k) are truncated

Blind inversion of Wiener systems iid criterion

- Output of the inversion (Hammerstein) structure:

$$y(t) = w * z(t) \qquad\qquad z(t) = g(x(t))$$

- **Y**(t) is spatially independent = the sequence {*y*(t)} is iid
- The mutual information for infinite dimensional stationary random vectors is defined from *entropy rates* (Cover, Thomas, John Wiley & Sons, 1991)

$$H(Y) = \lim_{T \to +\infty} \frac{1}{2T+1} H(y(-T),\dots,y(T))$$

$$I(Y) = \lim_{T \to \infty} \frac{1}{2T+1} \left\{ \sum_{t=-T}^{T} H(y(t)) - H(y(-T),\cdots,y(T)) \right\}$$

- *I*(Y) is always positive and vanishes if and only if *y(t)* is an iid sequence

Blind inversion of Wiener systems
Estimation equations (1/2)

- Estimation equations are then

$$\frac{\partial I(Y)}{\partial \boldsymbol{\theta}} = 0, \text{ with } \boldsymbol{\theta} \text{ is the parameter vector}$$

- *I(Y)* must be derived with respect to parameters of the linear part: *w,* and with respect to the nonlinear function: *g*
- We will use the relative gradient descent which provides equivariant algorithms.
- *Linear part*. Consider a small relative variation of w, in terms of a convolution by a small filter ε. The first order variation of *I(Y)*:

$\Delta I(Y) = -\left[\gamma_{y,\psi_y}(y) + \delta\right] * \varepsilon\,(0)$ with $\gamma_{y,\psi_y}(y)(\tau) = E\left[y(t-\tau)\,\psi_y(y(t))\right]$

with $\varepsilon = \mu\left\{\gamma_{y,\psi_y}(y) + \delta\right\}$ it leads to $\omega \leftarrow \omega + \mu\left\{\gamma_{y,\psi_y}(y) + \delta\right\} * \omega$

- With a Gaussian signal, $\gamma_{y,\psi_y}(y)$ reduces to second order statistics...

Blind inversion of Wiener systems Estimation equations (2/2)

- Using a nonparametric approach, the relative deviation:

$$g \leftarrow g + \varepsilon \circ g$$

the differential with respect to NL function g is:

$$\Delta I(y) = -\int \underbrace{E[\psi_y(y(\tau))\{w * \delta(x-v)\}(\tau) + \delta(x(\tau)-v)]}_{J(v)} \varepsilon(v)\,dv$$

- The gradient descent algorithm is then:

$$g \leftarrow g + \mu\{Q * J\} \circ g$$

provided than Q is any function satisfying:

$$\int J(v)\, Q * J(v)\, dv \geq 0$$

Blind inversion of Wiener systems
Practical issues

- Score function estimation based on kernel

$$\hat{p}_y(u) = \frac{1}{hT}\sum_{t=1}^{T} K\left(\frac{u-y(t)}{h}\right) \qquad \hat{\psi}_y(u) = \frac{\hat{p}'_y(u)}{\hat{p}_y(u)}$$

- Estimation of $\gamma_{y,\psi_y}(\tau)$

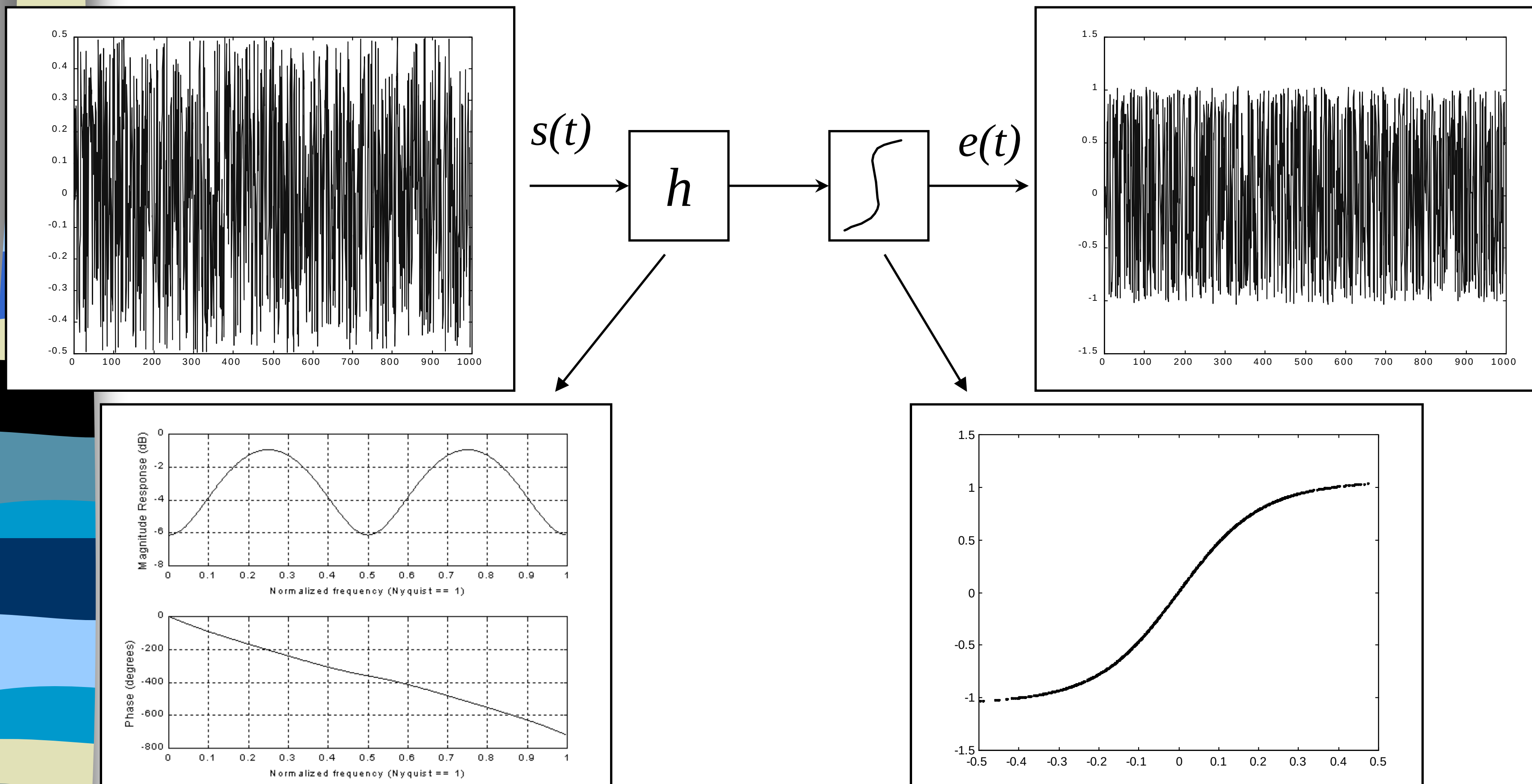
$$\hat{\gamma}_{y,\psi_y}(\tau) = \frac{1}{T}\sum_{\tau=1}^{T} y(t-\tau)\,\hat{\psi}_y(y(t))$$

- Estimation of *Q*J* : relative deviation of *g*

$$Q*J(v) = \frac{1}{T}\sum_{t=1}^{T} -Q'(v-x(t)) + \psi_y(y(t))\{w * Q(v-x)\}(t)$$

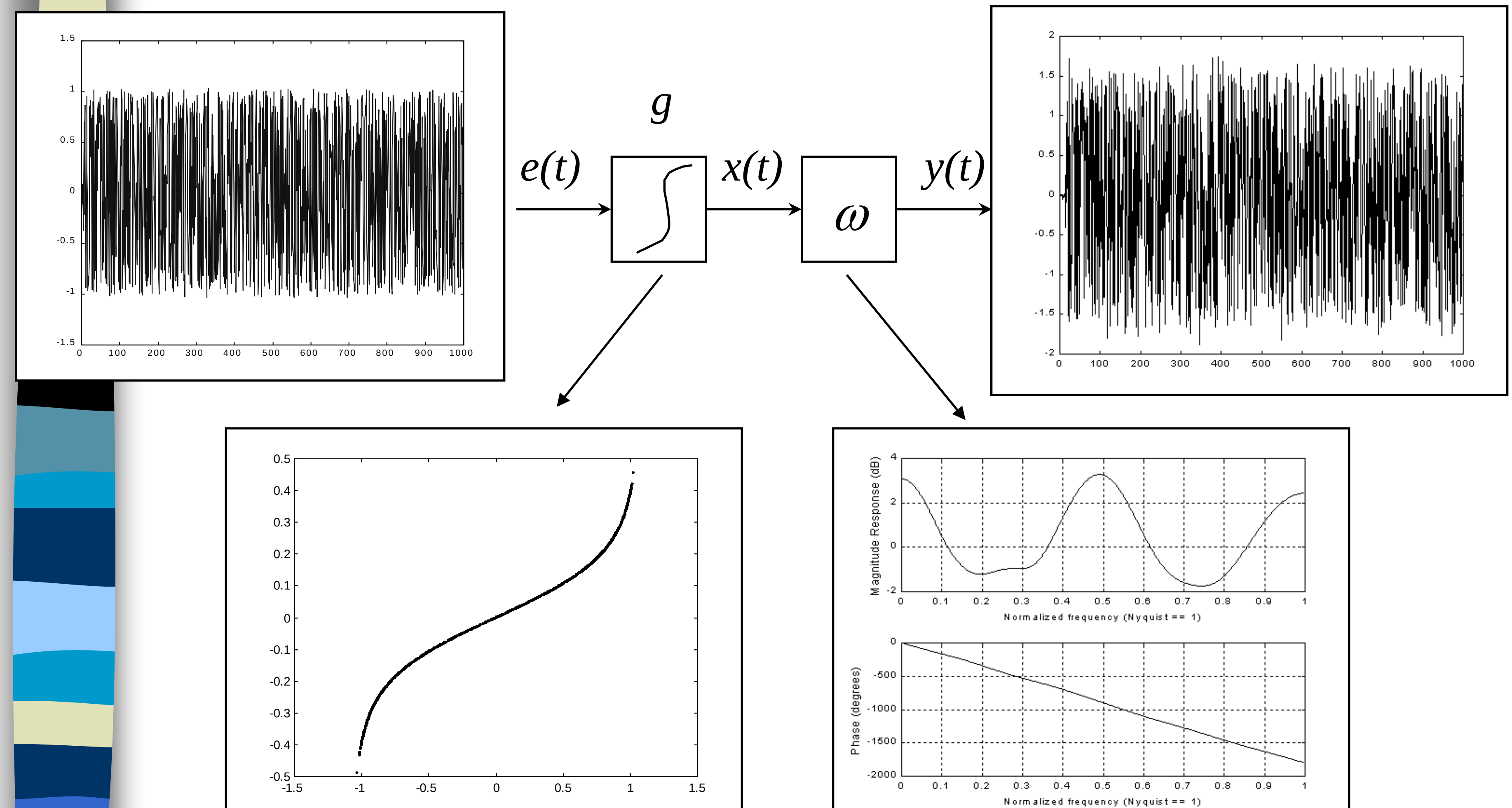
- A simple choice of Q is $Q(u) = \begin{cases} -u & u > 0 \\ 0 & \text{otherwise} \end{cases}$

Blind inversion of Wiener systems
Experiments (1/2)



$h = [0.082, 0, -0.1793, 0, 0.6579, 0, -0.1793, 0, -0.082]$ $\qquad f(u) = 0.5u + \tanh(5u)$

Blind inversion of Wiener systems Experiments (2/2)



Discussion and Perspectives

- Generally, independence is not sufficient for insuring separability
- Independence can insure separability in structured NL mixtures
- PNL mixtures are particular separable NL mixtures
- Deconvolution and Wiener system inversion are related to BSS:

  time independence (iid) is then related to spatial independence
- I(Y) can measure spatial independence or time independence (iid)
- Independence is powerful enough for *blindly* compensate strong NL distortions
  - in PNL multichannel systems: satellite antenna, sensors array, etc.
  - in Wiener systems (NL dynamic SISO)
- Extension to MIMO dynamic NL systems (Babaie-Zadeh, Jutten)
- Bilinear and multi-linear models (Hosseini, Deville, ICA 04)

7. Semi-blind approaches

Gaussian iid or non Gaussian

Discrete sources

Bounded and sparse sources

Time-frequency idea

Semi-blind approaches

- If there are more *a priori* informations, even very weak $\rightarrow$ Exploit them ! $\rightarrow$ Semi-Blind
- Advantages:
  - Improve the separation performance
  - Provide simpler algorithms
  - Can work when a blind solution is difficult
    - More sources than sensors
    - Separating Gaussian sources

Gaussian mixtures and 2^{nd} order methods

- Source separation not possible if sources are at the same time (Cardoso, ICA2001):
  – Gaussian
  – White (first “i” in “i.i.d”)
  – Stationary (“i.d.” in “i.i.d”)

- Any of these dropped $\Rightarrow$ Source separation is possible
  – Dropping Gaussianity $\Rightarrow$ iid non Gaussian : “Blind” (ICA) (Gaussian signals - except one - cannot be separated)
  – Dropping stationarity or whiteness $\Rightarrow$ Gaussian non iid: “Semi-Blind” (Gaussianity is not required, i.e. second-order statistics is enough, Gaussian signals can be separated)

Colored or Non-stationary sources

- A few advantages:
  - Independence is not required, only decorrelation
  - Only 2nd-order statistics
  - Able to separate Gaussian sources, but not only
  - Fast iterative algorithms for jointly diagonalizing matrices (JADE, SOBI, TDSEP, algo. of Yeredor, Pham, etc.)

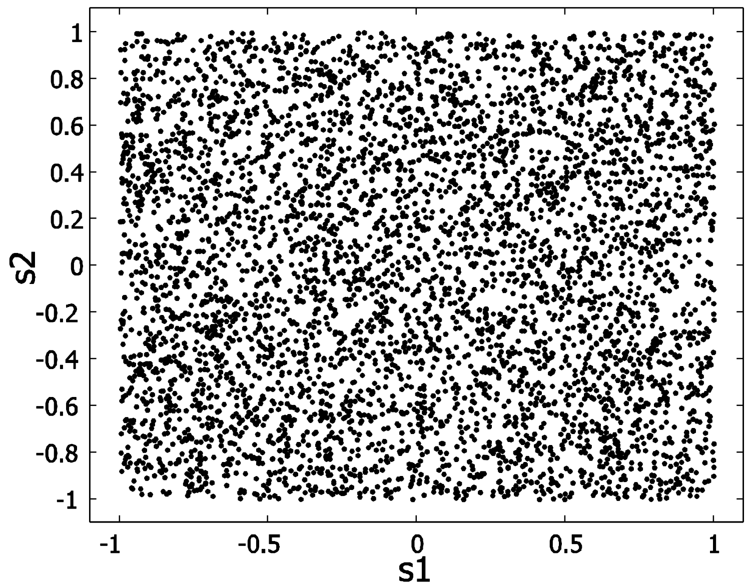
Examples of Semi-Blind approaches

- Geometrical approaches
  - Bounded sources
  - Discrete-valued sources
- Sparse sources
- Bayesian approaches
- Audio-Visual approaches

Geometric: Bounded Sources

- Independence $\Leftrightarrow p_{s1s2}(s_1,s_2)=p_{s1}(s_1)\,p_{s2}(s_2)$
- Bounded support for p_{s1} and $p_{s2} \Rightarrow$ rectangular support for p_{s1s2}
- $\Rightarrow$ scatter plot of sources is a rectangle

Example with uniformly distributed sources

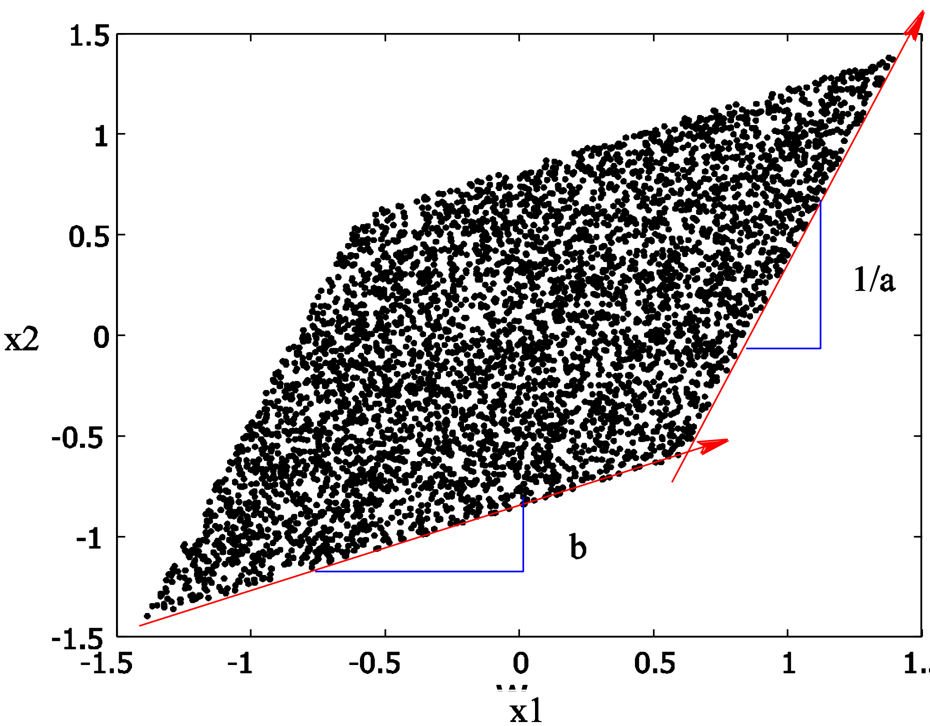


Bounded Sources (*cont.*)

- $\mathbf{x} = \mathbf{As}$ transforms this rectangle to a parallelogram
- Mixing matrix assumed:

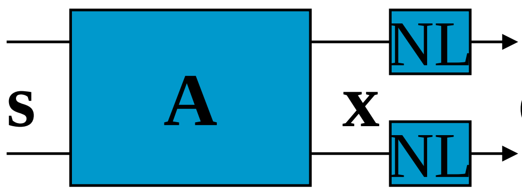
$$\mathbf{A} = \begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix}$$

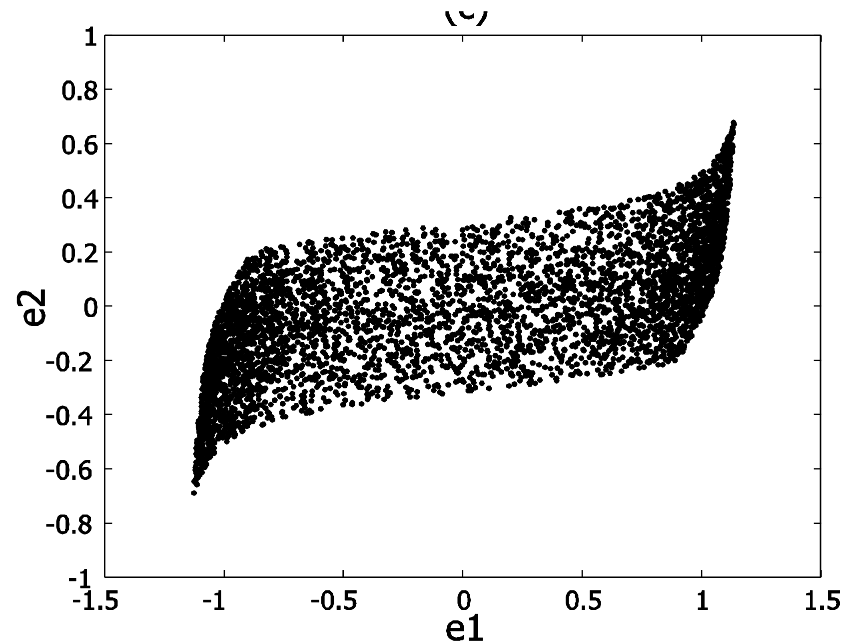
- Slopes of borders $\rightarrow$ $1/a$ and b $\rightarrow$ mixing matrix



Bounded Sources (*cont.*)

- Post Non-Linear (PNL) mixtures: linear mixtures and nonlinear sensors



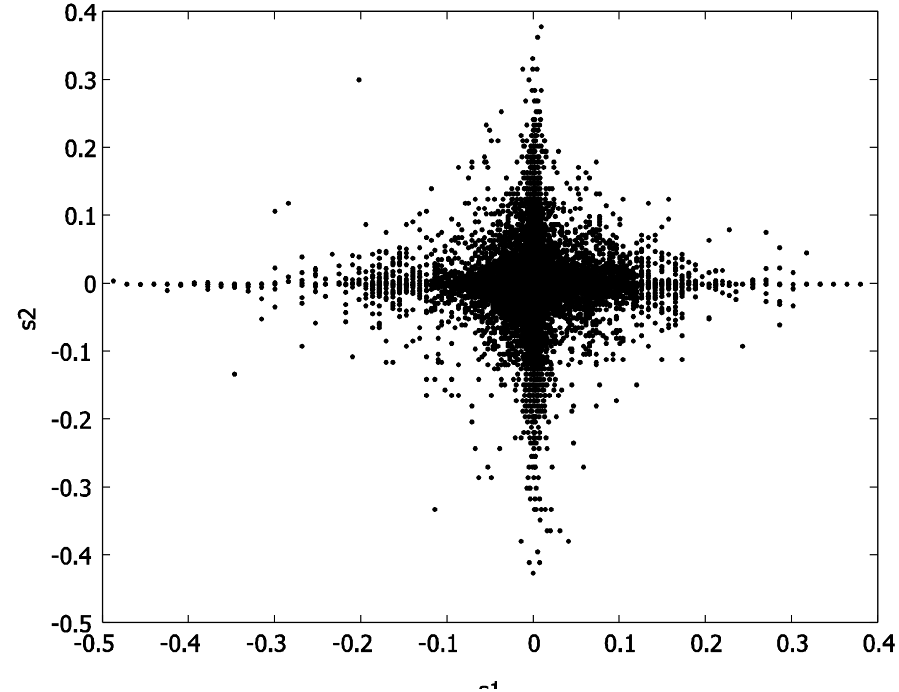


- Geometric: Transform the NL mixtures again to a parallelogram, and then separate the linear mixtures

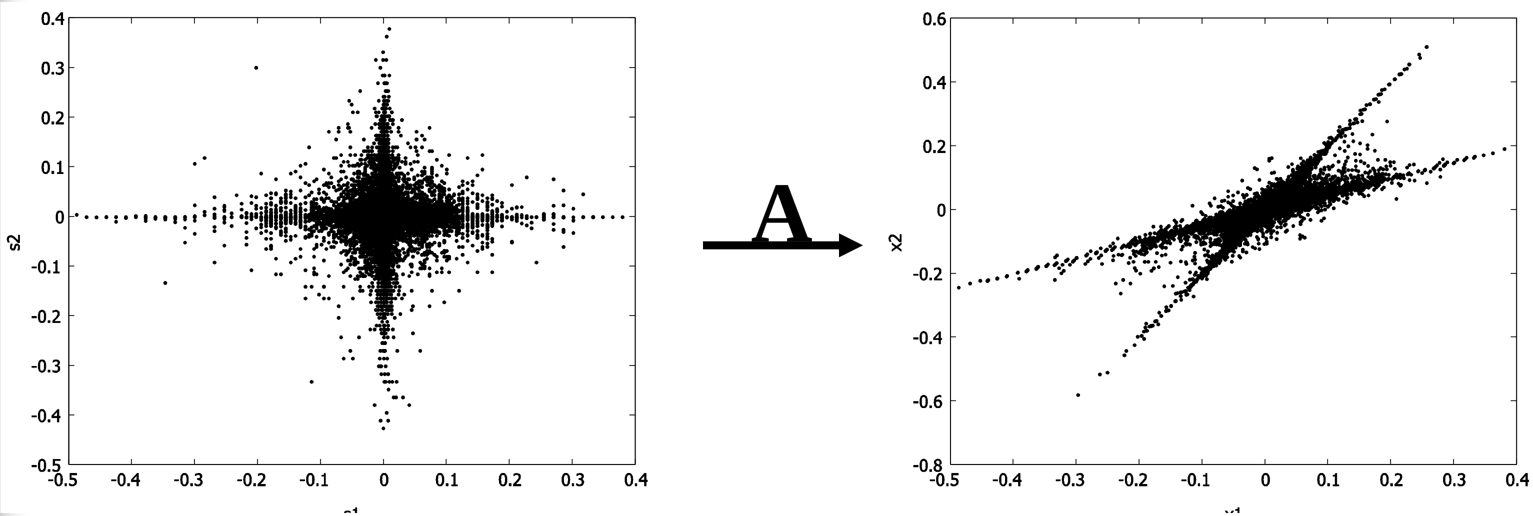
Geometric: Sparse sources

Like speech, ECG, EEG,...

- The rectangle is not well filled (requires lot of data sample).
- Source PDF's are concentrated around zero.
- Probability of having a point on the border of parallelogram is very small.

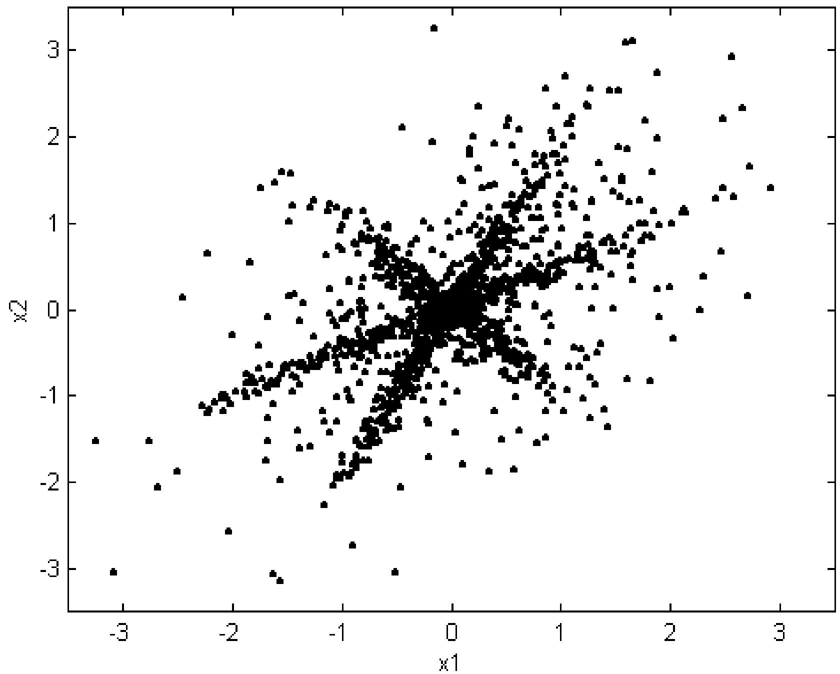


Sparse sources



- Geometrical approach: using “axes” instead of “borders”

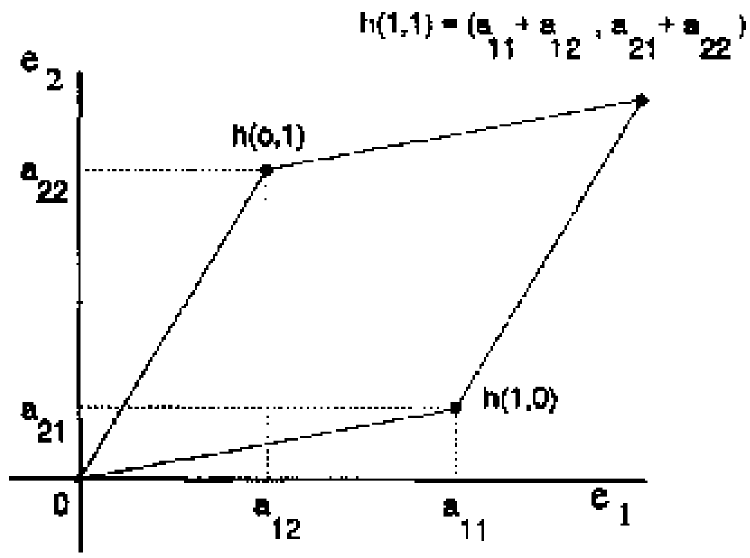
Sparse Sources

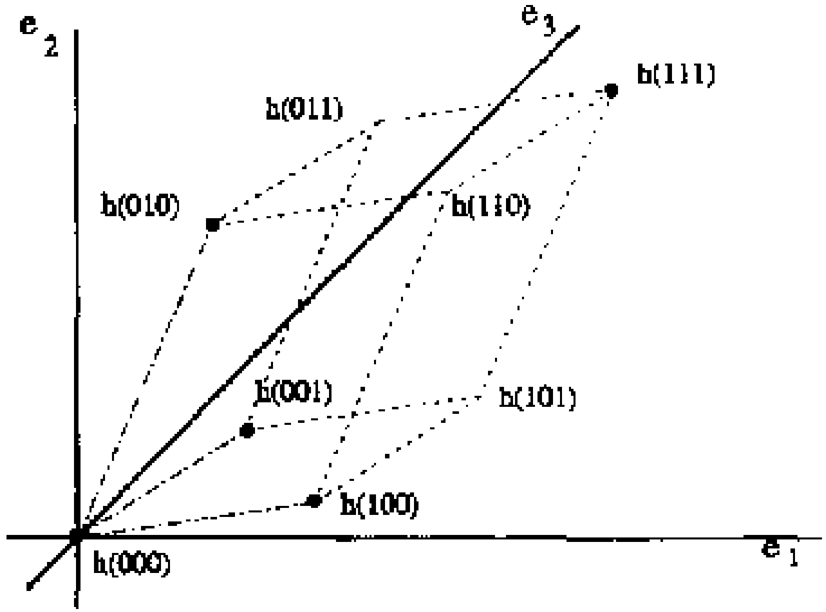


2 mixtures of 3 sources

- Possibility to separate more sources than sensors
- Identification of mixtures ≠ source separation

Discrete-Valued Sources





- (Belouchrani and Cardoso, 1994; Puntonet *et al*., 1995; Taleb and Jutten, 1999; Grellier and Comon, 1998)
- Other example of sparsity, usual in digital communications
- Possibility to separate more sources than sensors, robust to noise

Sparse Component Analysis

- Review paper (see Gribonval, ESANN'06)
- Main ideas
  - Initial source separation problem

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

  - Transform with a sparsifying transform T, which preserves linearity (e.g. wavelet transform, ST Fourier transform, etc.)

$$T(\mathbf{x}) = T(\mathbf{A}\mathbf{s}) = \mathbf{A}T(\mathbf{s}) \Leftrightarrow \tilde{\mathbf{x}} = \mathbf{A}\tilde{\mathbf{s}}$$

  - Solve the source separation problem in the sparse space

$$\Rightarrow \text{estimation } \hat{\tilde{\mathbf{s}}}$$

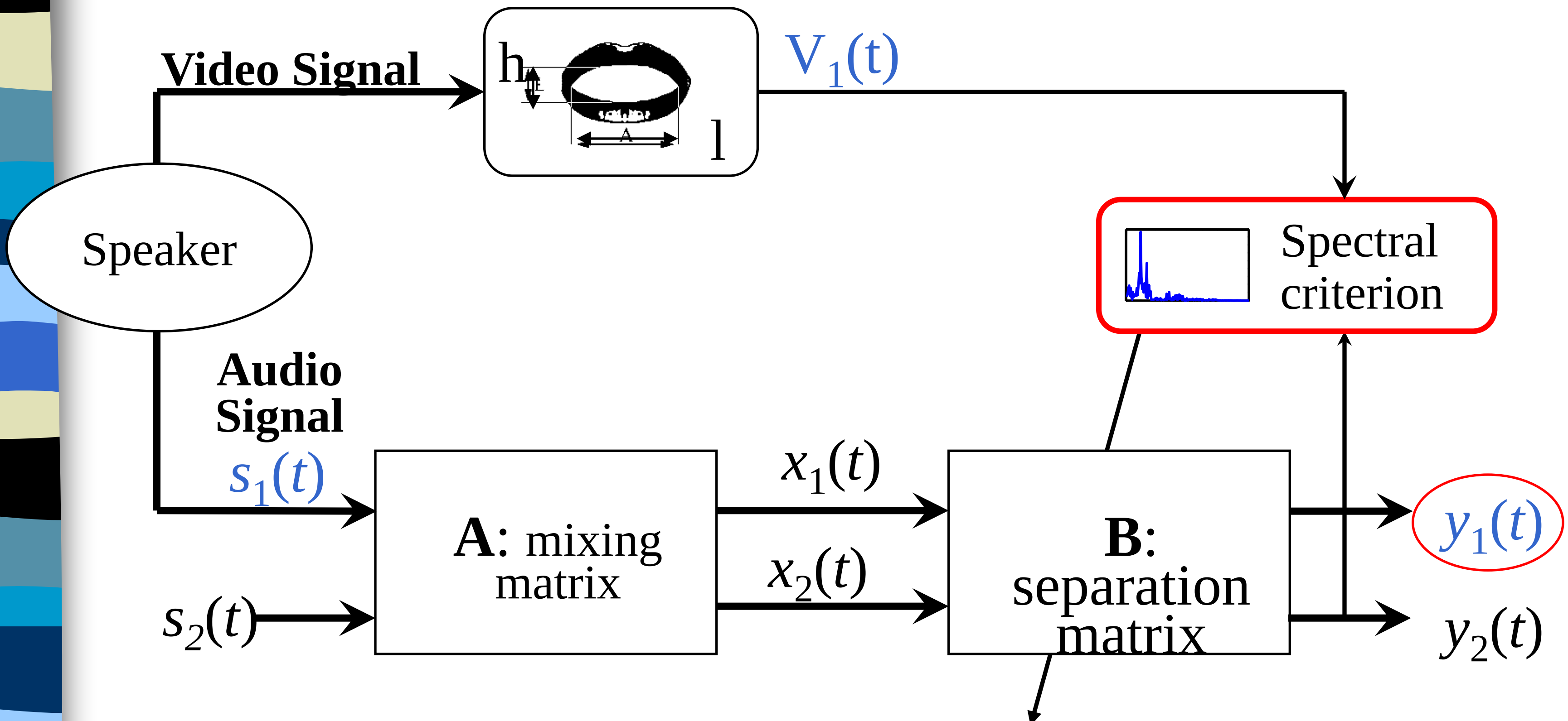
  - Come back to the initial space, with inverse of T

$$\hat{\tilde{\mathbf{s}}} \Leftrightarrow \hat{\mathbf{s}} = T^{-1}(\hat{\tilde{\mathbf{s}}})$$

Bayesian approaches

- Provide a general framework for modeling prior information:
  - source distribution,
  - time correlation,
  - additive noise,
  - …
- Can process more sources than sensors, and additive noise
- Main problem: time consuming, MCMC method !
- Review paper (Mohammad-Djafari, ESANN'06)

Audio-visual source extraction



Extraction of the source of interest

h, l, audio ⇒ p(spectrum/video, audio) ⇒ **B** estimated by ML

h, l ⇒ Voice activity detector ⇒ cancel permut. in convol. mixt.

(Sodoyer et al., IEEE ASSP, Rivet et al., IEEE ASSP)