Rao Blackwell approximation for fast MCMC implementation of Bayesian blind source separation problems.

Hichem SNOUSSI
and
Ali MOHAMMAD–DJAFARI

Laboratoire des Signaux et Systèmes
CNRS – SUPÉLEC – UPS

SUPÉLEC, Plateau de Moulon, 91192
Gif–sur–Yvette Cedex, France

Organization of the talk

- Source separation problem

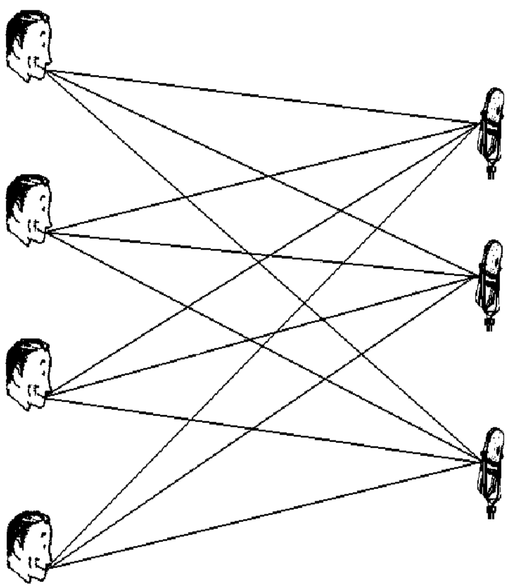
  $\longrightarrow$ Bayesian approach

- Sources modelling

  $\longrightarrow$ Hidden Markov Model

- EM algorithm.
- From SEM to RB-SEM.
- Application in satellite imaging.

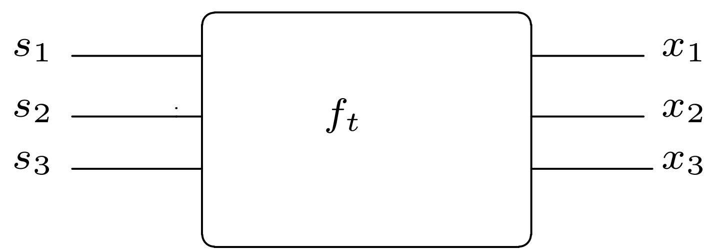
Problem description



Mixture of sounds

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

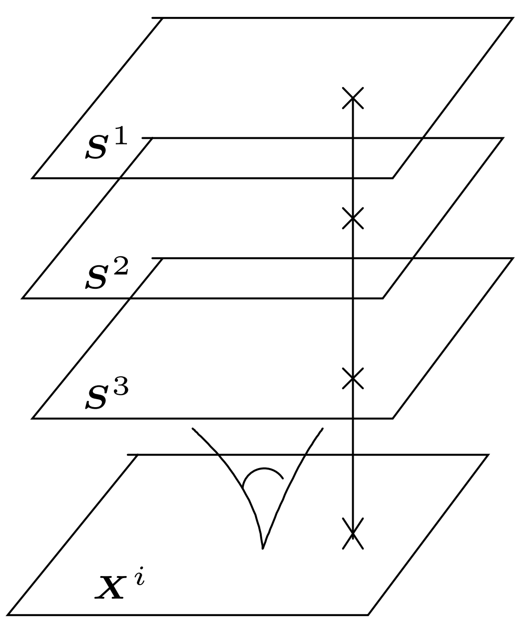
$\uparrow$



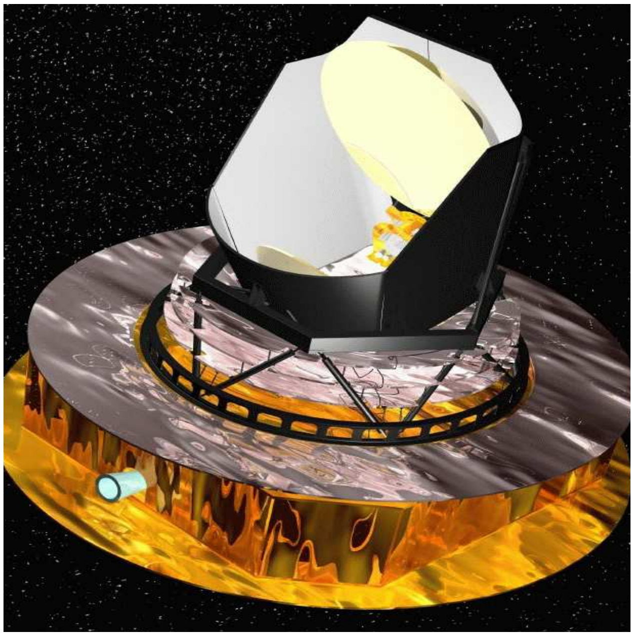
$\downarrow$

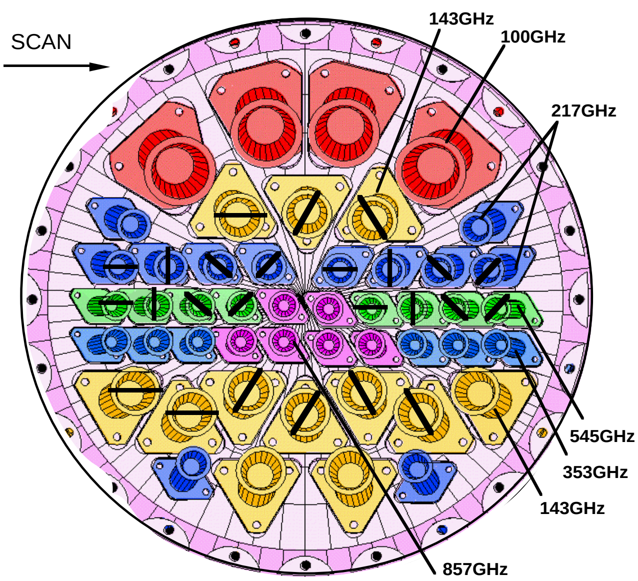
$$\boldsymbol{x}(i,j) = \boldsymbol{A}\boldsymbol{s}(i,j) + \boldsymbol{b}(i,j)$$

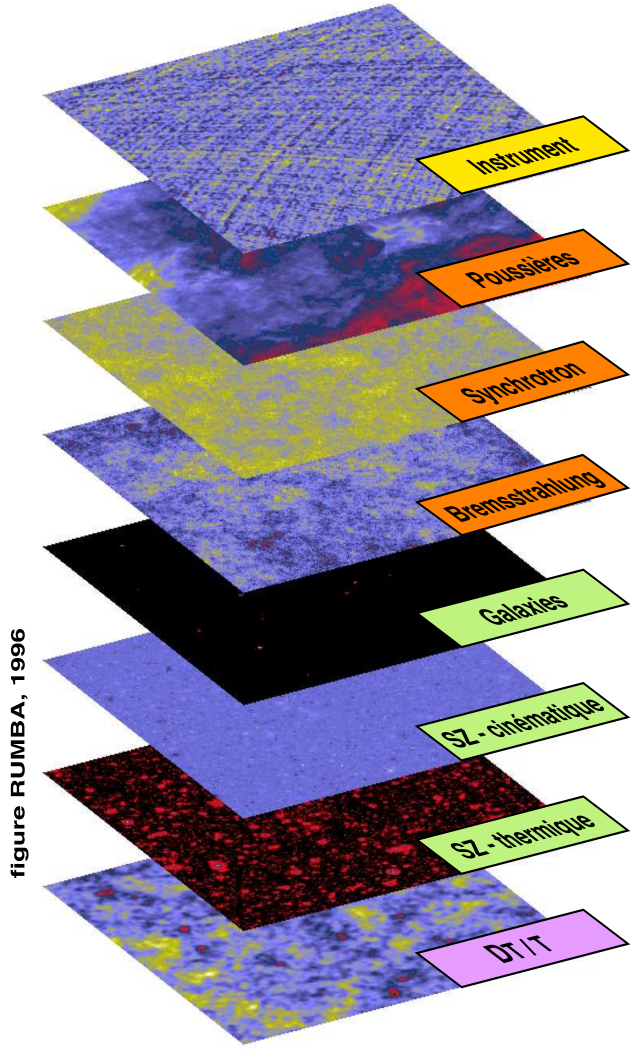
Mixture of images



Mixture of astrophysical emissions







$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{b}.$$

Bayesian approach

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{b}(t),\ t = 1..T.$$

- *a posteriori* distribution (Bayes rule) : [Mohammad-Djafari99, Knuth99]

$$\begin{aligned} \mathrm{p}(\boldsymbol{A} \,|\, \boldsymbol{x}_{1..T}, \boldsymbol{I}) &\propto \mathrm{p}(\boldsymbol{x}_{1..T} \,|\, \boldsymbol{A}, \boldsymbol{I})\mathrm{p}(\boldsymbol{A} \,|\, \boldsymbol{I}) \\ &\propto \int \mathrm{p}(\boldsymbol{x}_{1..T} \,|\, \boldsymbol{s}_{1..T}, \boldsymbol{A}) \underbrace{\mathrm{p}(\boldsymbol{s}_{1..T})}_{Choice\,?} d\,\boldsymbol{s}_{1..T} \\ &\quad \times\ \mathrm{p}(\boldsymbol{A} \,|\, \boldsymbol{I}) \end{aligned}$$

- Hidden **natural** structure :

  $\boldsymbol{x}_{1..T} \longrightarrow$ the incomplete data.

  $\boldsymbol{s}_{1..T} \longrightarrow$ the missing data.

- Relationship with independent component analysis (ICA) :

  - To take into account the **noise** in the model.

  - To exploit ***a priori*** information on the mixing coefficients $\longrightarrow$ regularization of ICA in the noiseless case.

- Choices of probabilities for the noise, the sources, the mixing matrix... ?

Sources modelling

- i.i.d non Gaussian [Gaeta90, Jutten91, Bermond00]
- correlated Gaussian [Belouchrani95]
- Non stationary Gausssians [Pham01]
- [Snoussi01d] : Mixture of distributions

$$\mathrm{p}(\boldsymbol{s}_{1..T}) = \sum_{\boldsymbol{z}_{1..T}} \mathrm{p}(\boldsymbol{s}_{1..T} \,|\, \boldsymbol{z}_{1..T}) P(\boldsymbol{z}_{1..T})$$

- double stochastic process :

Real sources	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{s}_3$	$\ldots$	$\boldsymbol{s}_T$
	$\uparrow$	$\uparrow$	$\uparrow$	$\ldots$	$\uparrow$
Hidden sources	$\boldsymbol{z}_1$	$\boldsymbol{z}_2$	$\boldsymbol{z}_3$	$\ldots$	$\boldsymbol{z}_T$

- Given the variables $\boldsymbol{z}_{1..T}$, the sources are temporally white :

$$\mathrm{p}(\boldsymbol{s}_{1..T} \,|\, \boldsymbol{z}_{1..T}) = \prod_{t=1}^{T} \mathrm{p}(\boldsymbol{s}_t \,|\, \boldsymbol{z}_t)$$

- The z_t take discrete values (classes) and are :

$\longrightarrow$ independent (mixture of Gaussians i.i.d),

$\longrightarrow$ or correlated with a Markovian structure (Markov chain 1-D, Markov field 2-D).

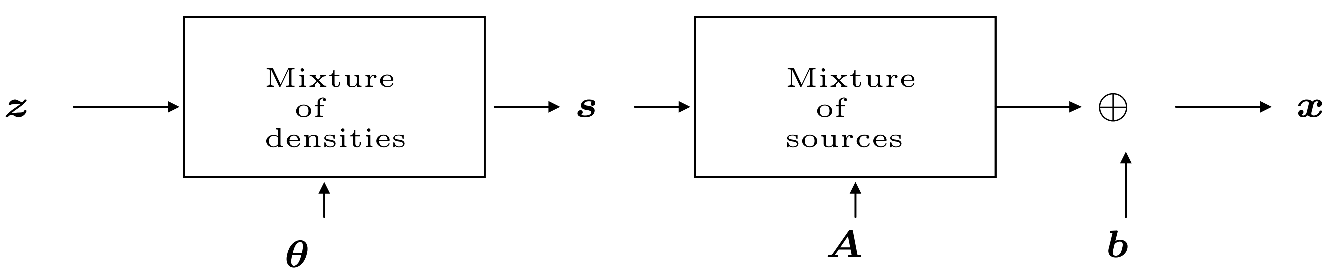
Double hidden structure

- The sources $(\boldsymbol{s}_1, ..., \boldsymbol{s}_T)$ are not directly observed :

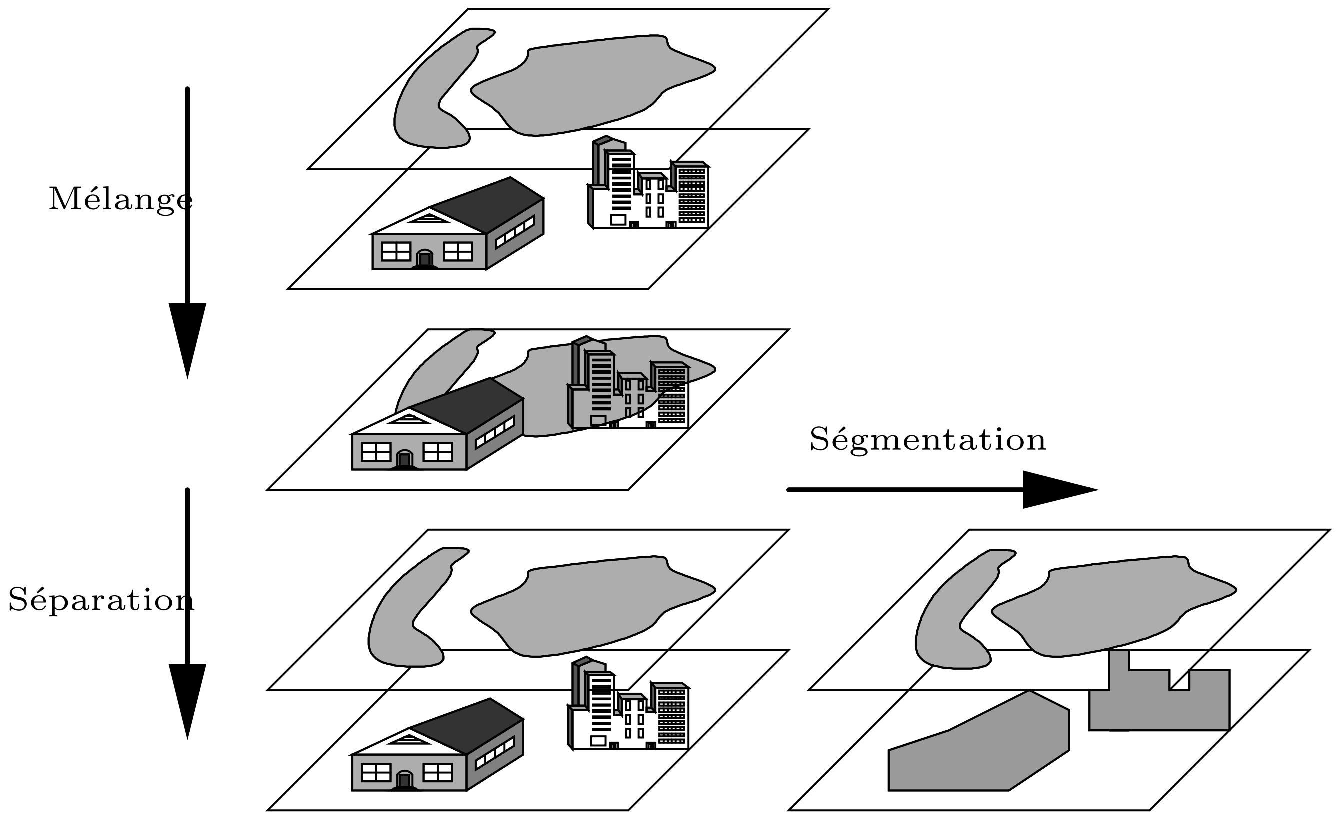
$$\begin{cases} \boldsymbol{x}_{1..T} = \boldsymbol{A}\,\boldsymbol{s}_{1..T} + \boldsymbol{b}_{1..T}, \quad t = 1..T, \\ \\ \mathrm{p}(\boldsymbol{s}_{1..T}) = \displaystyle\sum_{\boldsymbol{z}_{1..T}} \mathrm{p}(\boldsymbol{z}_{1..T})\mathrm{p}(\boldsymbol{s}_{1..T} \,|\, \boldsymbol{z}_{1..T}) \end{cases}$$

$\longrightarrow$ $(\boldsymbol{s}_1, ..., \boldsymbol{s}_T)$ form a second stage of hidden variables.

Mixed sources	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\dots$	$\boldsymbol{x}_T$
	$\uparrow$	$\uparrow$	$\uparrow$	$\dots$	$\uparrow$
Real sources	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{s}_3$	$\dots$	$\boldsymbol{s}_T$
	$\uparrow$	$\uparrow$	$\uparrow$	$\dots$	$\uparrow$
Hidden labels	z_1	z_2	z_3	$\dots$	z_T



Non stationarity



- Strong similarity between separation and segmentation : hidden variable problems $\longrightarrow$ algorithmic efficiency.

Interpretation of the separation criteria (given $\boldsymbol{z}_{1..T}$) :

$$\begin{aligned} \mathcal{J} &= \log\ \mathrm{p}(\boldsymbol{A} \,|\, \boldsymbol{x}_{1..T}, \boldsymbol{z}_{1..T}) \\ &= \underbrace{\sum_{z=1}^{K} \alpha_z D_{\mathcal{LK}}(\boldsymbol{R}_{xx} \,||\, \boldsymbol{A}\boldsymbol{R}_z \boldsymbol{A}^T + \boldsymbol{R}_\epsilon)}_{\textbf{Covariances matching [Pham01]}} + \underbrace{\log\ \mathrm{p}(\boldsymbol{A})}_{\textbf{regularization}} \end{aligned}$$

Algorithmic aspects

$$\boldsymbol{\eta} = (\boldsymbol{A}, \boldsymbol{R}_\epsilon, \boldsymbol{\theta})$$

- EM algorithm $\longrightarrow$ start at $\boldsymbol{\eta}^{(0)}$, iterate :

(i) E. (Expectation) $\longrightarrow$ functional computation :

$$\mathcal{Q}(\boldsymbol{\eta} \,|\, \boldsymbol{\eta}^{(k-1)}) = \mathop{\mathrm{E}}_{\boldsymbol{s},\boldsymbol{z}} \left[\log p(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{z} \,|\, \boldsymbol{\eta}) + \log p(\boldsymbol{\eta}) \,|\, \boldsymbol{x},\, \boldsymbol{\eta}^{(k-1)}\right]$$

(ii)M. (Maximization) $\longrightarrow$ functional maximization :

$$\boldsymbol{\eta}^{(k)} = \arg\max \mathcal{Q}(\boldsymbol{\eta} \,|\, \boldsymbol{\eta}^{(k-1)})$$

Computation of marginal probabilities $\mathrm{p}(\boldsymbol{z}_t \,|\, \boldsymbol{x}_{1..T}, \boldsymbol{\eta}^{(k-1)})$

- Cas 1-D : **[Snoussi02a]**
  - EM exact $\longrightarrow$ Baum Welsh procedure $[O([\prod_{j=1}^{n} K_j]^2\, T)]$.
  - Approximations of the EM :

$[O([\prod_{j=1}^{n} K_j]\, T)]$	$[O([\sum_{j=1}^{n} K_j]\, T)]$
Viterbi-EM, Gibbs-EM	Fast-Viterbi-EM, Fast-Gibbs-EM
reduction of the computation cost due to the markovian structure	reduction of the computation cost due to the spatial structure

Algorithmic aspects : 2-D case

❏ Computation of labels probabilities untractable.

❏ ⟶ Stochastic approximations of the EM

⟹ the **SEM** algorithm :

1. *a*- Simulate $\tilde{\boldsymbol{Z}} \sim \mathrm{p}(\boldsymbol{Z} \,|\, \boldsymbol{X}, \boldsymbol{\eta}^{(k-1)})$

   b- Simulate $\tilde{\boldsymbol{S}} \sim \mathrm{p}(\boldsymbol{S} \,|\, \boldsymbol{X}, \tilde{\boldsymbol{Z}}, \boldsymbol{\eta}^{(k-1)})$

   c- Compute the functional :

$$\tilde{\mathcal{Q}}(\boldsymbol{\eta} \,|\, \boldsymbol{\eta}^{(k-1)}) = \log\, \mathrm{p}(\boldsymbol{X}, \tilde{\boldsymbol{S}}, \tilde{\boldsymbol{Z}} \,|\, \boldsymbol{\eta}) + \log \mathrm{p}(\boldsymbol{\eta})$$

⟶ Unbiased estimator of $\mathcal{Q}(\boldsymbol{\eta} \,|\, \boldsymbol{\eta}^{(k-1)})$ (the EM functional).

2. $\boldsymbol{\eta}^{(k)} = \arg\max_{\boldsymbol{\eta}} \tilde{\mathcal{Q}}(\boldsymbol{\eta} \,|\, \boldsymbol{\eta}^{(k-1)})$.

❏ $(\boldsymbol{\eta}^{(k)})_{k\in}$ is a Markov chain.

❏ Under some regularity conditions [Nielsen97], the estimator $\boldsymbol{\eta}^{(k)}$ (in its **stationary** regime) is asymptotically consistent.

From SEM to RB-SEM

- RB-SEM (Rao Blackwell SEM) :

  1. Simulate M samples $\boldsymbol{Z}^{(m)}$ (M images $\boldsymbol{Z}$) according to $\mathrm{p}(\boldsymbol{Z} \,|\, \boldsymbol{X}, \boldsymbol{\eta}^{(k-1)})$

  2. Compute the functional :

$$\tilde{\mathcal{Q}}(\boldsymbol{\eta} \,|\, \boldsymbol{\eta}^{(k-1)}) = \frac{1}{M} \sum_{m} \mathop{\mathrm{E}}_{s} \left[\log \mathrm{p}(\boldsymbol{X}, \boldsymbol{S}, \boldsymbol{Z}^{(m)} \,|\, \boldsymbol{\eta})\right] + \log \mathrm{p}(\boldsymbol{\eta})$$

$\longrightarrow$ Empirical sum on $\boldsymbol{Z}$ and exact integration with respect to $\boldsymbol{S}$.

$\longrightarrow$ unbiased estimator of $\mathcal{Q}(\boldsymbol{\eta} \,|\, \boldsymbol{\eta}^{(k-1)})$

  3. $\boldsymbol{\eta}^{(k)} = \arg\max_{\boldsymbol{\eta}} \tilde{\mathcal{Q}}(\boldsymbol{\eta} \,|\, \boldsymbol{\eta}^{(k-1)})$.

$$M \begin{cases} \nearrow \quad \infty & \longrightarrow \quad \text{MCEM} \\ \searrow \quad < \infty & \longrightarrow \quad \text{RB-SEM} \end{cases}$$

From SEM to RB-SEM

• RB-SEM more efficient than the SEM algorithm due to the Rao Blackwell property when integrating with respect to the sources :

$$Var(\tilde{\eta}_n^{RB-SEM}) \leq Var(\tilde{\eta}_n^{SEM})$$

• RB-SEM less time consuming than SEM.

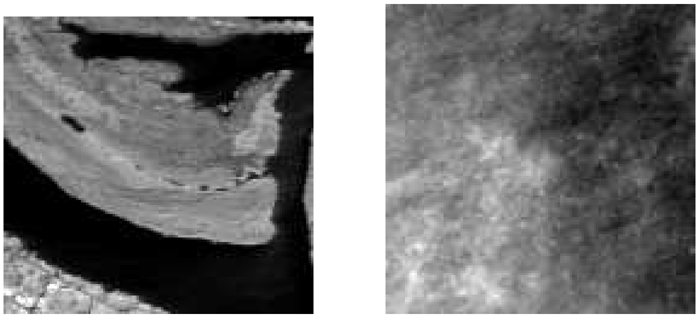
• RB-SEM is easily implemented for general incomplete data structure problems.

• The introduction of a second hidden set of random variables :

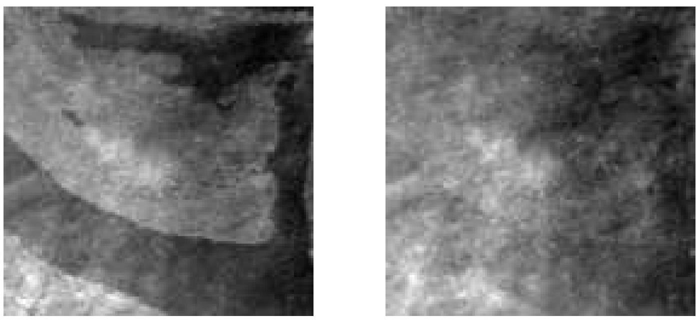
$\longrightarrow$ good alternative to non parametrics.

$\longrightarrow$ more robustness to stochastic versions of EM type algorithm.

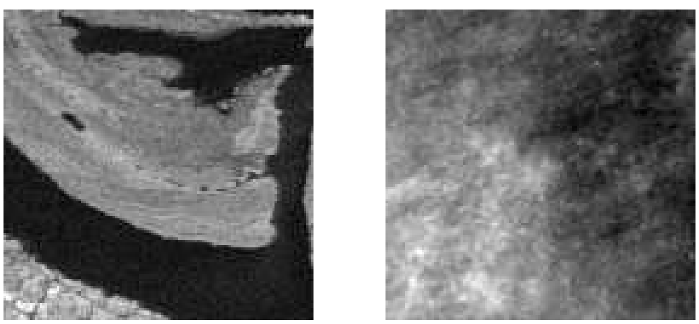
Joint separation and segmentation



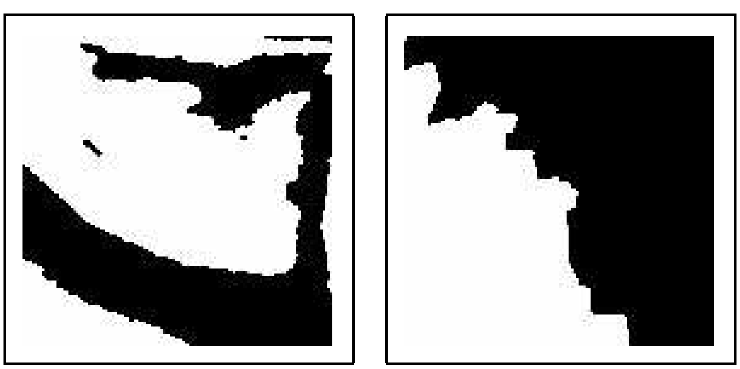
Original sources



Mixed sources



Separated sources



Segmented sources