

# ANALYSE EN FACTEURS INDÉPENDANTS POUR LE DIAGNOSTIC D'UN COMPOSANT DE L'INFRASTRUCTURE FERROVIAIRE DANS UN CADRE SEMI-SUPERVISÉ

Soutenance de Stage - M2 IAD

Nicolas CHEIFETZ

Encadrante INRETS :  
Latifa OUKHELLOU



Encadrant LIP6:  
Patrick GALLINARI



14 septembre 2009

- 1 Introduction
  - Problématique
- 2 Modèle de mélanges et Algorithme EM
  - Modèle de mélanges
  - Algorithme EM
- 3 Analyse en Facteurs Indépendants
  - Principe
  - Cadre du problème
  - Application : circuits de voie
  - Deux a priori
  - Extension à un nombre variable de sources
- 4 Conclusion

# RAPPEL DE PLAN

- 1 Introduction
  - Problématique
- 2 Modèle de mélanges et Algorithme EM
  - Modèle de mélanges
  - Algorithme EM
- 3 Analyse en Facteurs Indépendants
  - Principe
  - Cadre du problème
  - Application : circuits de voie
  - Deux a priori
  - Extension à un nombre variable de sources
- 4 Conclusion

# PRÉAMBULE

## Problématique industrielle

Diagnostic d'un système complexe de taille variable :

- détecter et localiser le(s) sous-système(s) défectueux
- estimer la gravité du (des) défaut(s)

Conséquence : accroître la fiabilité et la durabilité du système sans diminuer sa disponibilité

⇒ Résolution par Reconnaissance des formes statistique

Modèle d'apprentissage étudié : Analyse en Facteurs Indépendants

Deux a priori :

- seule une partie des données apprises est labellisée
- connaissance de certaines dépendances spatiales entre sous-systèmes

# PRÉAMBULE

## Problématique industrielle

Diagnostic d'un système complexe de taille variable :

- détecter et localiser le(s) sous-système(s) défectueux
- estimer la gravité du (des) défaut(s)

Conséquence : accroître la fiabilité et la durabilité du système sans diminuer sa disponibilité

⇒ Résolution par Reconnaissance des formes statistique

Modèle d'apprentissage étudié : **Analyse en Facteurs Indépendants**

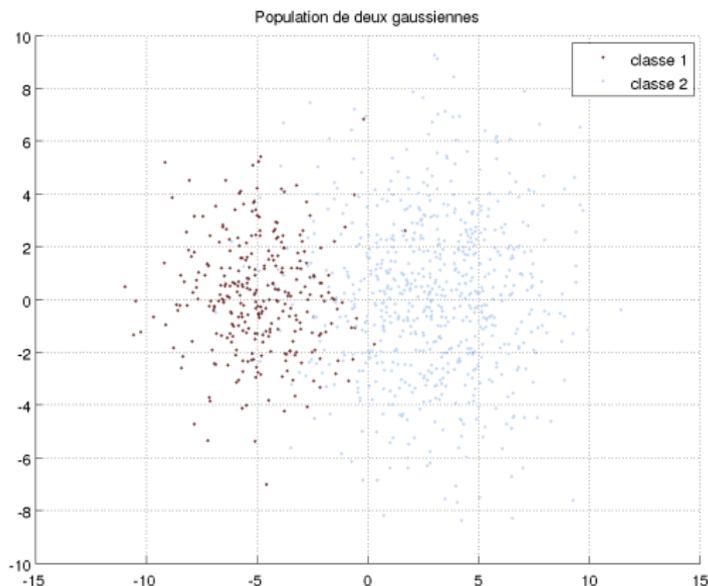
Deux a priori :

- seule une partie des données apprises est labellisée
- connaissance de certaines dépendances spatiales entre sous-systèmes

# RAPPEL DE PLAN

- 1 Introduction
  - Problématique
- 2 **Modèle de mélanges et Algorithme EM**
  - Modèle de mélanges
  - Algorithme EM
- 3 Analyse en Facteurs Indépendants
  - Principe
  - Cadre du problème
  - Application : circuits de voie
  - Deux a priori
  - Extension à un nombre variable de sources
- 4 Conclusion

# MODÈLE DE MÉLANGES, UNE APPROCHE GÉNÉRATIVE



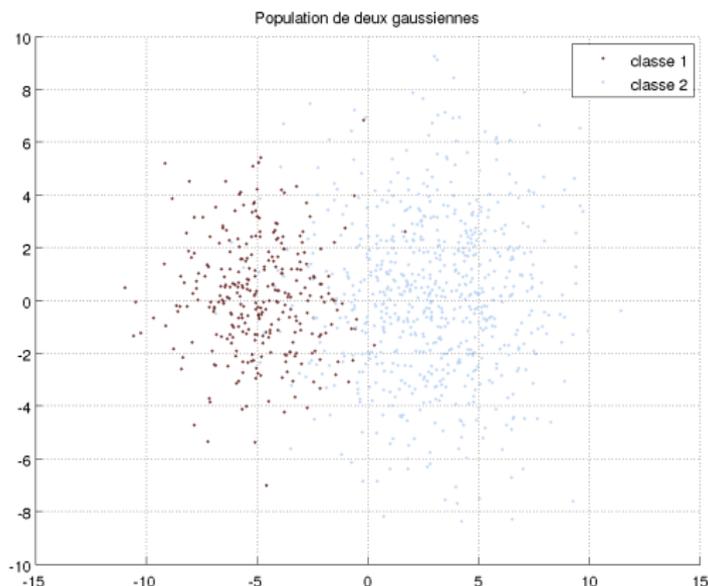
Proportion a priori

$$\pi_k = p(Y_i = C_k)$$

Densité gaussienne de chaque classe

$$f_k(x_i; \underbrace{\mu_k, \Sigma_k}_{\theta_k}) = \mathcal{N}(x_i; \theta_k)$$

# MODÈLE DE MÉLANGES, UNE APPROCHE GÉNÉRATIVE



## Proportion a priori

$$\pi_k = p(Y_i = C_k)$$

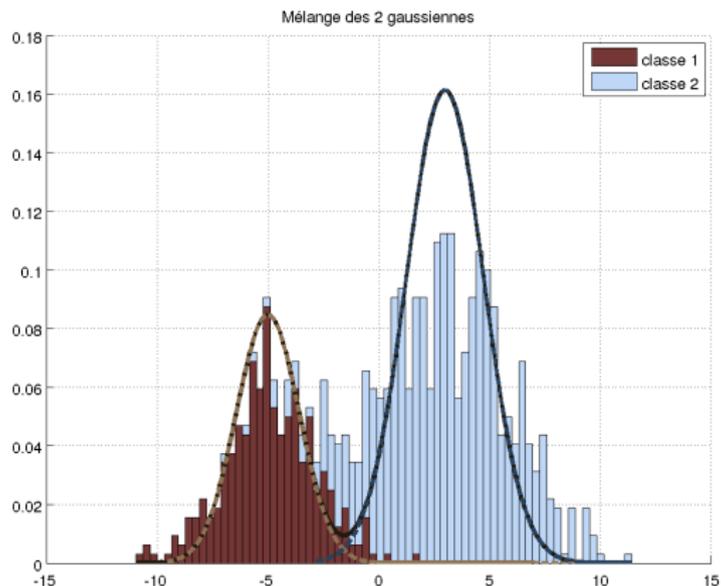
## Densité gaussienne de chaque classe

$$f_k(x_i; \underbrace{\mu_k, \Sigma_k}_{\theta_k}) = \mathcal{N}(x_i; \theta_k)$$

## Paramètre global

$$\Phi = (\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K)$$

# MODÈLE DE MÉLANGES, UNE APPROCHE GÉNÉRATIVE



## Proportion a priori

$$\pi_k = p(Y_i = C_k)$$

## Densité gaussienne de chaque classe

$$f_k(x_i; \underbrace{\mu_k, \Sigma_k}_{\theta_k}) = \mathcal{N}(x_i; \theta_k)$$

## Paramètre global

$$\Phi = (\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K)$$

## Densité d'un mélange gaussien

$$\forall i \in \{1, \dots, N\}, \quad f(x_i; \Phi) = \sum_{k=1}^K \pi_k \cdot \mathcal{N}(x_i; \theta_k)$$

# CRITÈRE DE L'APPROCHE GÉNÉRATIVE

- *Hypothèse* :  $x_i$  est i.i.d. ,  $\forall i \in \{1, \dots, N\}$
- Vraisemblance des données observées :

$$L(\Phi; \mathbb{X}) = \prod_{i=1}^N f(x_i; \Phi)$$

## Critère de Log-vraisemblance

$$\mathcal{L}(\Phi; \mathbb{X}) = \sum_{i=1}^N \log \left( \sum_{k=1}^K \pi_k \cdot \mathcal{N}(x_i; \theta_k) \right)$$

# CRITÈRE DE L'APPROCHE GÉNÉRATIVE

- *Hypothèse* :  $x_i$  est i.i.d. ,  $\forall i \in \{1, \dots, N\}$
- Vraisemblance des données observées :

$$L(\Phi; \mathbb{X}) = \prod_{i=1}^N f(x_i; \Phi)$$

## Critère de Log-vraisemblance

$$\mathcal{L}(\Phi; \mathbb{X}) = \sum_{i=1}^N \log \left( \sum_{k=1}^K \pi_k \cdot \mathcal{N}(x_i; \theta_k) \right)$$

## Objectif

$$\Phi^* = \arg \max_{\Phi} \mathcal{L}(\Phi; \mathbb{X})$$

# ALGORITHME EM (EXPECTATION-MAXIMISATION)

- Astuce de l'algorithme EM :

$$\mathcal{L}(\Phi; \mathbb{X}) = \mathbb{E}_{\mathcal{X}, \mathcal{Y}} (\mathcal{L}(\Phi; X) | X = \mathbb{X}, \Phi^{(q)})$$

# ALGORITHME EM (EXPECTATION-MAXIMISATION)

- Astuce de l'algorithme EM :

$$\begin{aligned}
 \mathcal{L}(\Phi; \mathbb{X}) &= \mathbb{E}_{X, Y} (\mathcal{L}(\Phi; X) | X = \mathbb{X}, \Phi^{(q)}) \\
 &= \underbrace{\mathbb{E}_{X, Y} (\log p(X, Y; \Phi) | \mathbb{X}, \Phi^{(q)})}_{Q(\Phi, \Phi^{(q)})} - \underbrace{\mathbb{E}_{X, Y} (\log p(X, Y | X; \Phi) | \mathbb{X}, \Phi^{(q)})}_{H(\Phi, \Phi^{(q)})}
 \end{aligned}$$

# ALGORITHME EM (EXPECTATION-MAXIMISATION)

- Astuce de l'algorithme EM :

$$\begin{aligned} \mathcal{L}(\Phi; \mathbb{X}) &= \mathbb{E}_{X,Y} (\mathcal{L}(\Phi; X) | X = \mathbb{X}, \Phi^{(q)}) \\ &= \underbrace{\mathbb{E}_{X,Y} (\log p(X, Y; \Phi) | \mathbb{X}, \Phi^{(q)})}_{Q(\Phi, \Phi^{(q)})} - \underbrace{\mathbb{E}_{X,Y} (\log p(X, Y | X; \Phi) | \mathbb{X}, \Phi^{(q)})}_{H(\Phi, \Phi^{(q)})} \end{aligned}$$

- Algorithme itératif en deux étapes :

- 1 **Etape E** (Estimation) : Calcul de la fonction  $Q$
- 2 **Etape M** (Maximisation) : Calcul du paramètre  $\Phi^{(q+1)}$

$$\Phi^{(q+1)} = \arg \max_{\Phi} Q(\Phi, \Phi^{(q)})$$

# ALGORITHME EM (EXPECTATION-MAXIMISATION)

- Astuce de l'algorithme EM :

$$\begin{aligned} \mathcal{L}(\Phi; \mathbb{X}) &= \mathbb{E}_{X,Y} (\mathcal{L}(\Phi; X) | X = \mathbb{X}, \Phi^{(q)}) \\ &= \underbrace{\mathbb{E}_{X,Y} (\log p(X, Y; \Phi) | \mathbb{X}, \Phi^{(q)})}_{Q(\Phi, \Phi^{(q)})} - \underbrace{\mathbb{E}_{X,Y} (\log p(X, Y | X; \Phi) | \mathbb{X}, \Phi^{(q)})}_{H(\Phi, \Phi^{(q)})} \end{aligned}$$

- Algorithme itératif en deux étapes :

- 1 **Etape E** (Estimation) : Calcul de la fonction  $Q$
  - 2 **Etape M** (Maximisation) : Calcul du paramètre  $\Phi^{(q+1)}$
- $$\Phi^{(q+1)} = \arg \max_{\Phi} Q(\Phi, \Phi^{(q)})$$

## Propriétés

- augmentation de la vraisemblance à chaque itération
- convergence vers un maximum local
- dépendance de l'initialisation

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

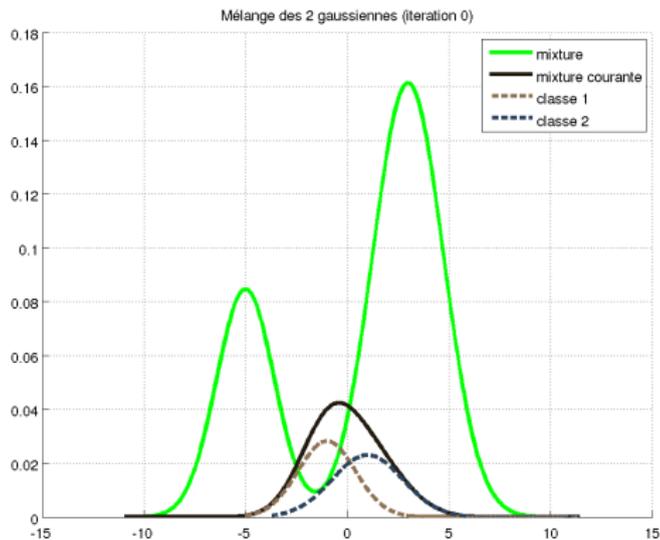
- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

- Ici,  $\epsilon = 10^{-5}$  :

- ▶  $\pi = (0.3, 0.7)$
- ▶  $\mu = (-5, 3)$
- ▶  $\Sigma = (2, 3)$



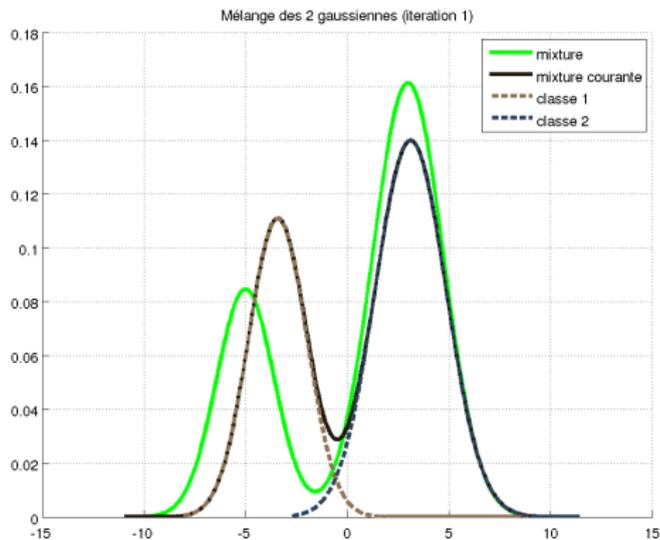
| Itération | $\mathcal{L}$ | Critère |
|-----------|---------------|---------|
| 0         | -6610.9       | —       |

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

- Ici,  $\epsilon = 10^{-5}$  :

- ▶  $\pi = (0.3, 0.7)$
- ▶  $\mu = (-5, 3)$
- ▶  $\Sigma = (2, 3)$



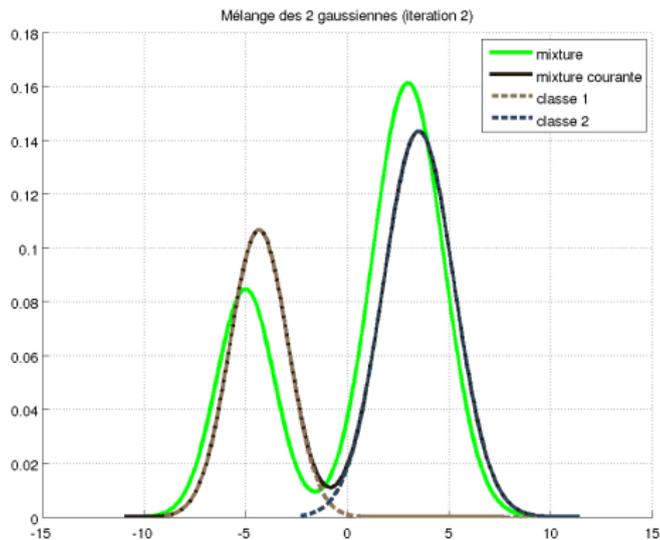
| Itération | $\mathcal{L}$ | Critère  |
|-----------|---------------|----------|
| 0         | -6610.9       | —        |
| 1         | -3947.0       | 0.402957 |

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

- Ici,  $\epsilon = 10^{-5}$  :

- ▶  $\pi = (0.3, 0.7)$
- ▶  $\mu = (-5, 3)$
- ▶  $\Sigma = (2, 3)$



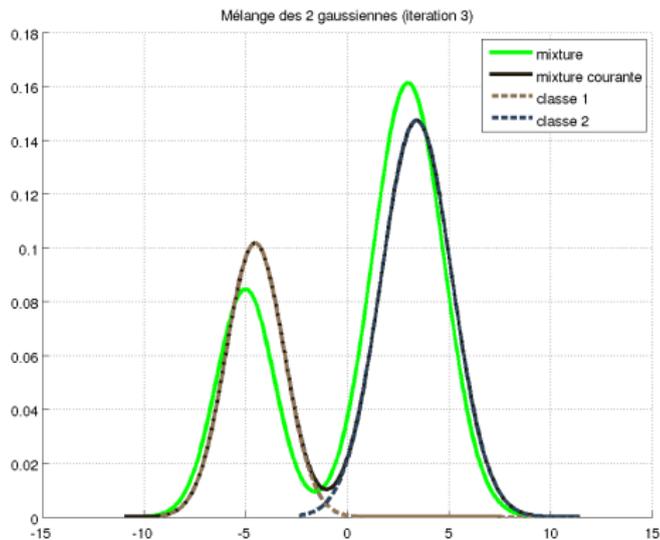
| Itération | $\mathcal{L}$ | Critère  |
|-----------|---------------|----------|
| 0         | -6610.9       | —        |
| 1         | -3947.0       | 0.402957 |
| 2         | -3126.0       | 0.029938 |

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

- Ici,  $\epsilon = 10^{-5}$  :

- ▶  $\pi = (0.3, 0.7)$
- ▶  $\mu = (-5, 3)$
- ▶  $\Sigma = (2, 3)$



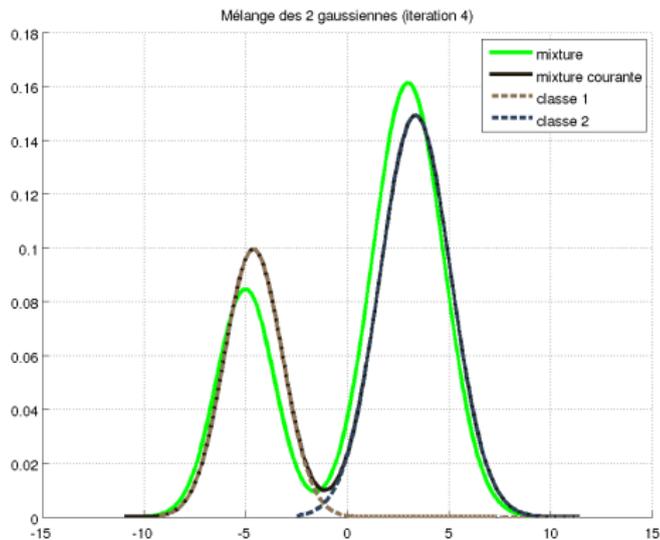
| Itération | $\mathcal{L}$ | Critère  |
|-----------|---------------|----------|
| 0         | -6610.9       | —        |
| 1         | -3947.0       | 0.402957 |
| 2         | -3126.0       | 0.029938 |
| 3         | -3093.1       | 0.001656 |

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

- Ici,  $\epsilon = 10^{-5}$  :

- ▶  $\pi = (0.3, 0.7)$
- ▶  $\mu = (-5, 3)$
- ▶  $\Sigma = (2, 3)$



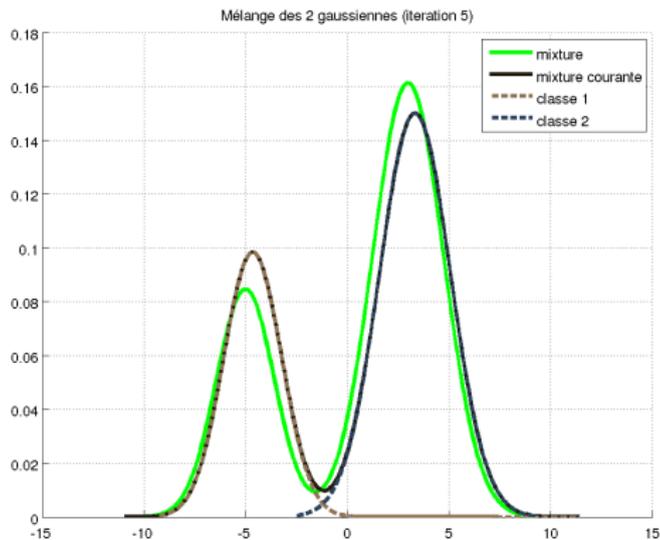
| Itération | $\mathcal{L}$ | Critère  |
|-----------|---------------|----------|
| 0         | -6610.9       | —        |
| 1         | -3947.0       | 0.402957 |
| 2         | -3126.0       | 0.029938 |
| 3         | -3093.1       | 0.001656 |
| 4         | -3086.6       | 0.000354 |

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

- Ici,  $\epsilon = 10^{-5}$  :

- ▶  $\pi = (0.3, 0.7)$
- ▶  $\mu = (-5, 3)$
- ▶  $\Sigma = (2, 3)$



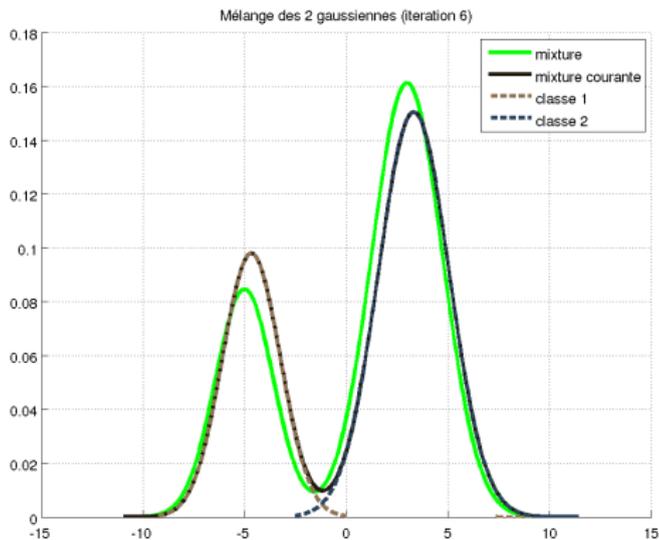
| Itération | $\mathcal{L}$ | Critère  |
|-----------|---------------|----------|
| 0         | -6610.9       | —        |
| 1         | -3947.0       | 0.402957 |
| 2         | -3126.0       | 0.029938 |
| 3         | -3093.1       | 0.001656 |
| 4         | -3086.6       | 0.000354 |
| 5         | -3085.1       | 0.000070 |

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

- Ici,  $\epsilon = 10^{-5}$  :

- ▶  $\pi = (0.3, 0.7)$
- ▶  $\mu = (-5, 3)$
- ▶  $\Sigma = (2, 3)$



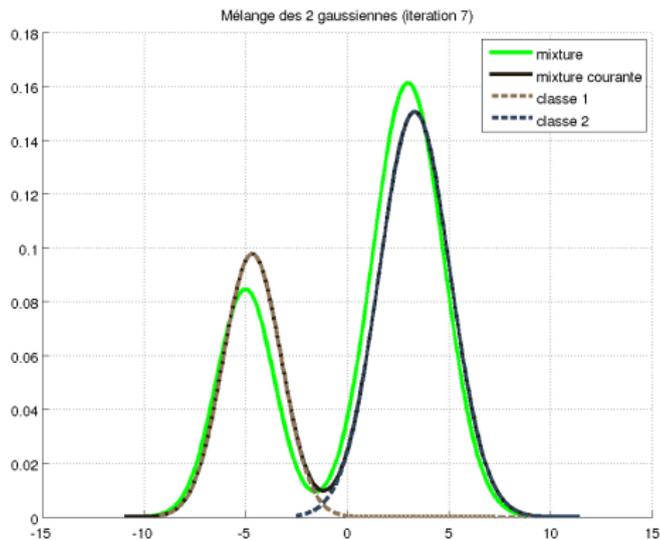
| Itération | $\mathcal{L}$ | Critère  |
|-----------|---------------|----------|
| 0         | -6610.9       | —        |
| 1         | -3947.0       | 0.402957 |
| 2         | -3126.0       | 0.029938 |
| 3         | -3093.1       | 0.001656 |
| 4         | -3086.6       | 0.000354 |
| 5         | -3085.1       | 0.000070 |
| 6         | -3084.6       | 0.000014 |

# EXEMPLE SUR UN MÉLANGE DE 2 GAUSSIENNES

- Critère de convergence : 
$$\left| \frac{\mathcal{L}(\Phi^{(q+1)}; \mathbb{X}) - \mathcal{L}(\Phi^{(q)}; \mathbb{X})}{\mathcal{L}(\Phi^{(q)}; \mathbb{X})} \right| < \epsilon$$

- Ici,  $\epsilon = 10^{-5}$  :

- ▶  $\pi = (0.3, 0.7)$
- ▶  $\mu = (-5, 3)$
- ▶  $\Sigma = (2, 3)$



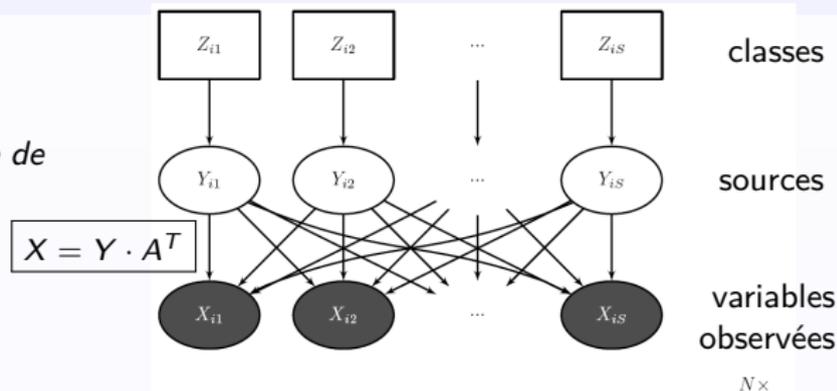
| Itération | $\mathcal{L}$ | Critère         |
|-----------|---------------|-----------------|
| 0         | -6610.9       | —               |
| 1         | -3947.0       | 0.402957        |
| 2         | -3126.0       | 0.029938        |
| 3         | -3093.1       | 0.001656        |
| 4         | -3086.6       | 0.000354        |
| 5         | -3085.1       | 0.000070        |
| 6         | -3084.6       | 0.000014        |
| 7         | -3084.4       | <b>0.000003</b> |

# RAPPEL DE PLAN

- 1 Introduction
  - Problématique
  
- 2 Modèle de mélanges et Algorithme EM
  - Modèle de mélanges
  - Algorithme EM
  
- 3 Analyse en Facteurs Indépendants
  - Principe
  - Cadre du problème
  - Application : circuits de voie
  - Deux a priori
  - Extension à un nombre variable de sources
  
- 4 Conclusion

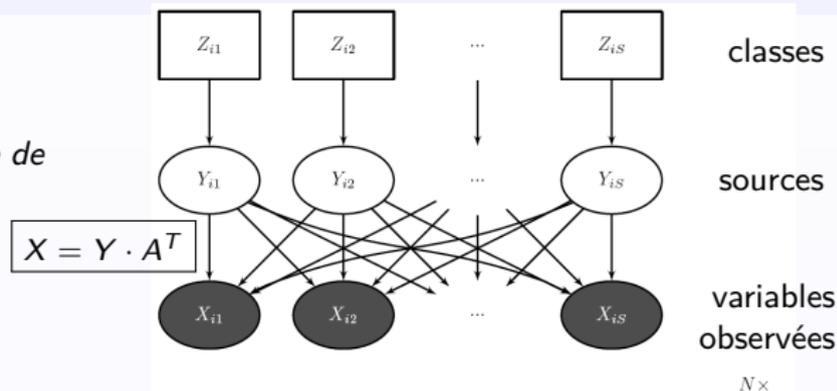
## ANALYSE EN FACTEURS INDÉPENDANTS [IFA] SANS BRUIT

FIGURE: *Modèle graphique de l'IFA sans bruit.*



## ANALYSE EN FACTEURS INDÉPENDANTS [IFA] SANS BRUIT

FIGURE: *Modèle graphique de l'IFA sans bruit.*

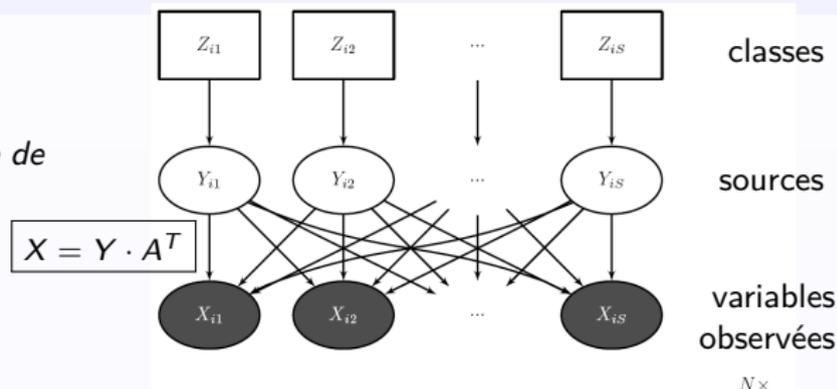


## Hypothèses du modèle

- variables observées = combinaison linéaire de sources
- indépendance des variables latentes
- indépendance conditionnelle des observations % sources
- densité de chaque source = mélange gaussien

## ANALYSE EN FACTEURS INDÉPENDANTS [IFA] SANS BRUIT

FIGURE: *Modèle graphique de l'IFA sans bruit.*



## Hypothèses du modèle

- variables observées = combinaison linéaire de sources
- indépendance des variables latentes
- indépendance conditionnelle des observations % sources
- densité de chaque source = mélange gaussien
- Optimisation par un algorithme **GEM** (Generalized EM) :
  - ▶ paramétrisation des sources ( $\pi_1, \dots, \pi_S, \theta_1, \dots, \theta_S$ )
  - ▶ montée de gradient pour la matrice de mixage  $A$

# CADRE DU PROBLÈME

- Notations :

- ▶ observations :  $\mathbb{X} = \{x_1, \dots, x_i, \dots, x_N\}$ , où  $x_i \in \mathbb{R}^S$
- ▶ variables latentes continues :  $\mathbb{Y} = \{y_1, \dots, y_i, \dots, y_N\}$ , où  $y_i \in \mathbb{R}^S$
- ▶ variables latentes discrètes :  $\mathbb{Z} = \{z_1, \dots, z_i, \dots, z_N\}$ , où  $z_i \in \mathbb{C}^S$

- Objectifs :

- ▶ estimation du paramètre global :  $\Phi = (A, \pi_1, \dots, \pi_S, \theta_1, \dots, \theta_S)$
- ▶ segmentation des observations / clustering

# CADRE DU PROBLÈME

- Notations :

- ▶ observations :  $\mathbb{X} = \{x_1, \dots, x_i, \dots, x_N\}$ , où  $x_i \in \mathbb{R}^S$
- ▶ variables latentes continues :  $\mathbb{Y} = \{y_1, \dots, y_i, \dots, y_N\}$ , où  $y_i \in \mathbb{R}^S$
- ▶ variables latentes discrètes :  $\mathbb{Z} = \{z_1, \dots, z_i, \dots, z_N\}$ , où  $z_i \in \mathcal{C}^S$

- Objectifs :

- ▶ estimation du paramètre global :  $\Phi = (A, \pi_1, \dots, \pi_S, \theta_1, \dots, \theta_S)$
- ▶ segmentation des observations / clustering

## Log-vraisemblance des données observées

$$\mathcal{L}(\Phi; \mathbb{X}) = -N \cdot \log |\det(A)| + \sum_{i=1}^N \sum_{s=1}^S \log \left( \sum_{k_s=1}^{K_s} \pi_{k_s} \cdot \mathcal{N}(y_{is}; \theta_{k_s}) \right)$$

avec  $y_{is} = x_i \cdot ({}^T A^{-1})_{.s}$

# CIRCUITS DE VOIE FERROVIAIRE

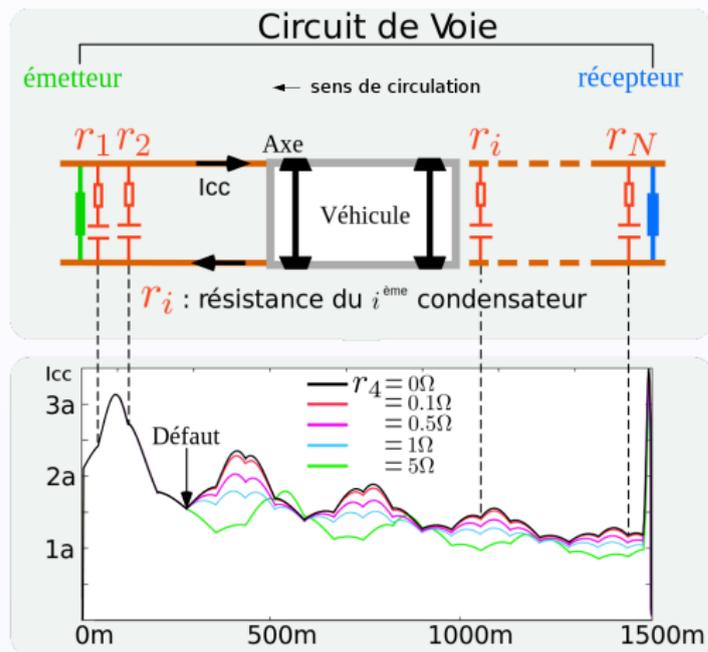
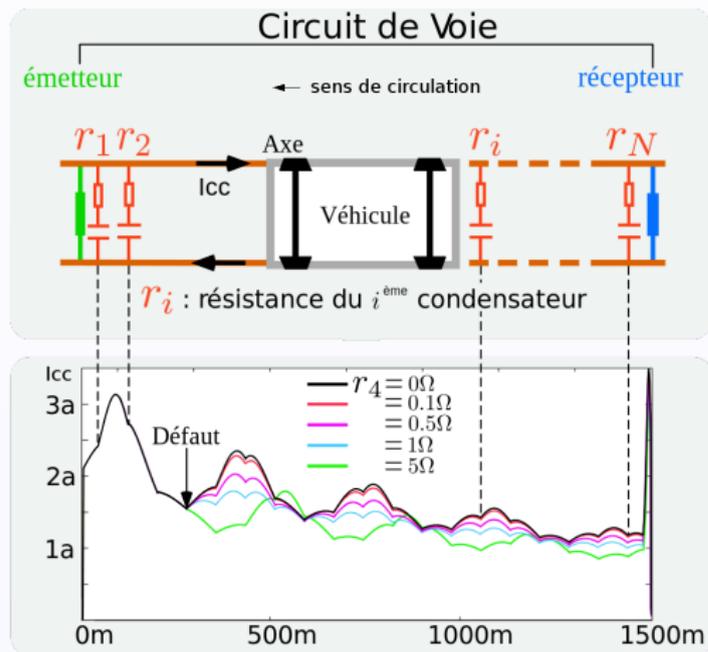


FIGURE: Schéma d'un circuit de voie ferroviaire.

## Circuits de Voie

- permet de détecter la présence d'un train

# CIRCUITS DE VOIE FERROVIAIRE



**FIGURE:** Schéma d'un circuit de voie ferroviaire.

## Circuits de Voie

- permet de détecter la présence d'un train

## Diagnostic d'un CdV

- forme du signal  $I_{cc}$

# CIRCUITS DE VOIE FERROVIAIRE

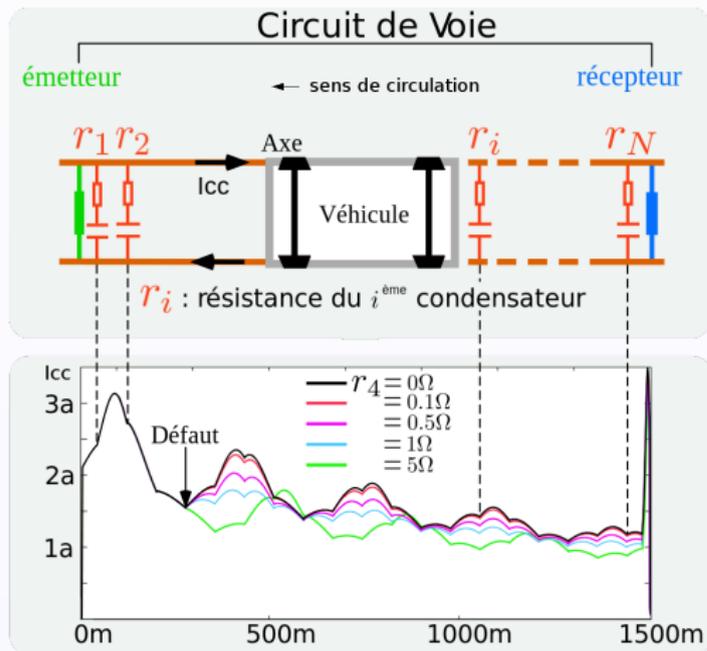


FIGURE: Schéma d'un circuit de voie ferroviaire.

## Circuits de Voie

- permet de détecter la présence d'un train

## Diagnostic d'un CdV

- forme du signal  $I_{cc}$
- diagnostic des condensateurs de compensation

# CIRCUITS DE VOIE FERROVIAIRE

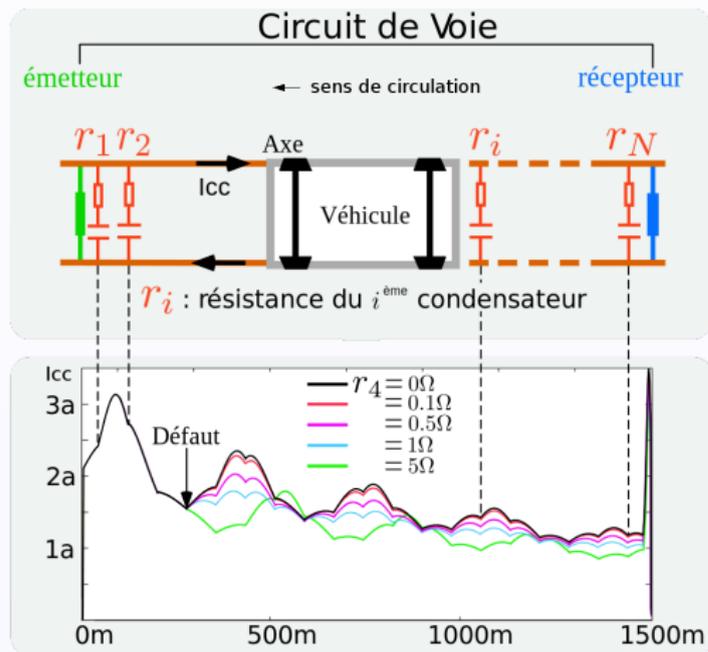


FIGURE: Schéma d'un circuit de voie ferroviaire.

## Circuits de Voie

- permet de détecter la présence d'un train

## Diagnostic d'un CdV

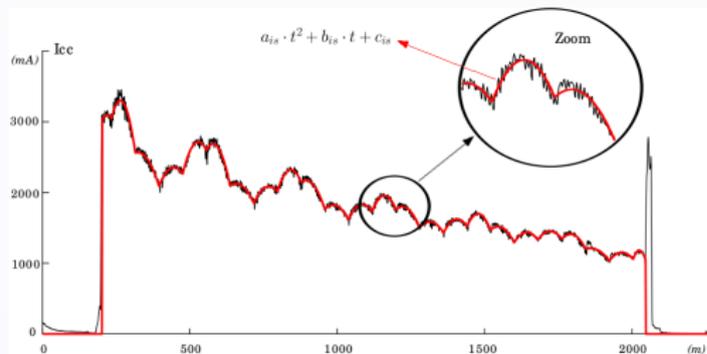
- forme du signal  $I_{cc}$
- diagnostic des condensateurs de compensation

## Notations :

|          |          |                        |
|----------|----------|------------------------|
| $N$      | $\equiv$ | # d'exemples de CdVs   |
| $Y_{is}$ | $\equiv$ | capacité d'un condo    |
| $Z_{is}$ | $\equiv$ | état d'un condensateur |
| $L$      | $\equiv$ | # de condos par CdV    |
| $p$      | $\equiv$ | # de coefs par arche   |

avec  $S = p \cdot L$

# CIRCUITS DE VOIE FERROVIAIRE



**FIGURE:** *Extraction des données observées à partir du signal  $I_{cc}$ .*

## Diagnostic d'un CdV

- forme du signal  $I_{cc}$
- diagnostic des condensateurs de compensation

Notations :

|          |          |                        |
|----------|----------|------------------------|
| $N$      | $\equiv$ | # d'exemples de CdVs   |
| $Y_{is}$ | $\equiv$ | capacité d'un condo    |
| $Z_{is}$ | $\equiv$ | état d'un condensateur |
| $L$      | $\equiv$ | # de condos par CdV    |
| $p$      | $\equiv$ | # de coefs par arche   |

avec  $S = p \cdot L$

# CIRCUITS DE VOIE FERROVIAIRE

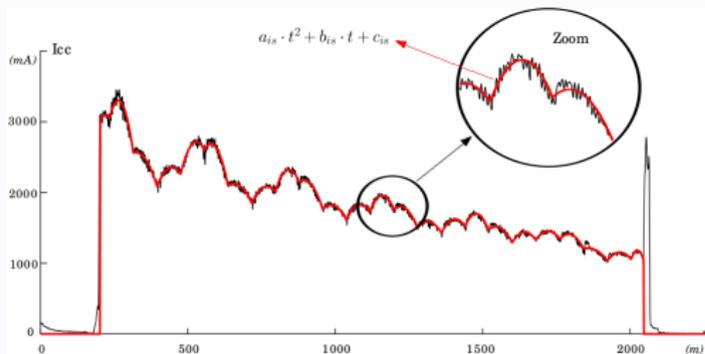


FIGURE: Extraction des données observées à partir du signal  $I_{cc}$ .

## Diagnostic d'un CdV

- forme du signal  $I_{cc}$
- diagnostic des condensateurs de compensation

Notations :

|          |          |                        |
|----------|----------|------------------------|
| $N$      | $\equiv$ | # d'exemples de CdVs   |
| $Y_{is}$ | $\equiv$ | capacité d'un condo    |
| $Z_{is}$ | $\equiv$ | état d'un condensateur |
| $L$      | $\equiv$ | # de condos par CdV    |
| $p$      | $\equiv$ | # de coefs par arche   |

avec  $S = p \cdot L$

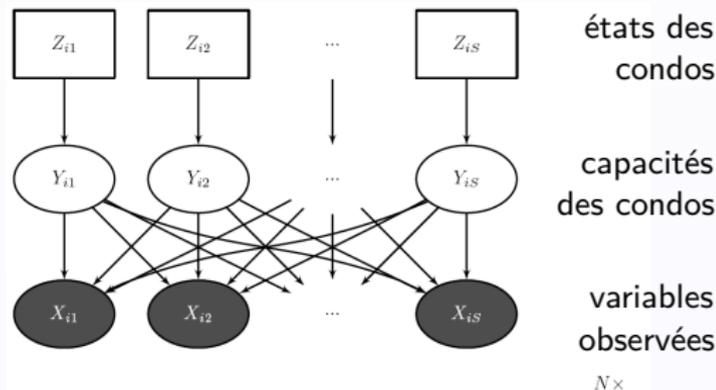
## Une observation

$$x_{is} = (b_{is}, c_{is})$$

# DEUX A PRIORI

- a priori sur les dépendances entre variables observées et sources

FIGURE: *Modèle graphique de l'IFA bruit appliquée aux CdVs sans contraintes spatiales.*

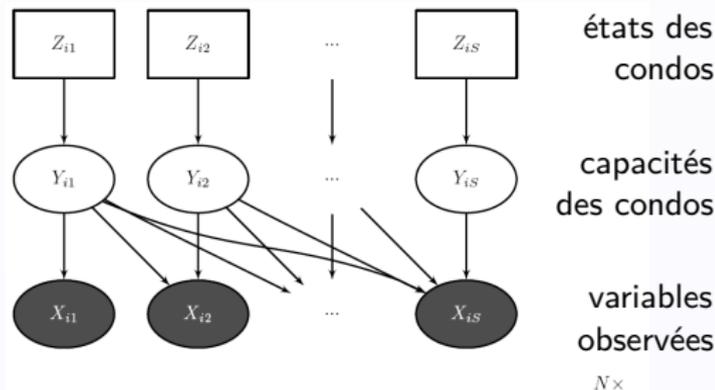


# DEUX A PRIORI

- a priori sur les dépendances entre variables observées et sources

- ▶ masque sur la matrice de mixage  $A$
- ▶ évite les permutations entre les sources

FIGURE: *Modèle graphique de l'IFA sans bruit appliquée aux CdVs avec contraintes spatiales.*

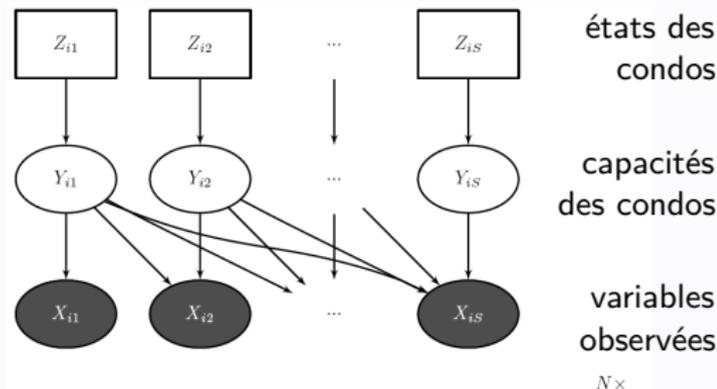


# DEUX A PRIORI

- a priori sur les dépendances entre variables observées et sources

- ▶ masque sur la matrice de mixage  $A$
- ▶ évite les permutations entre les sources

FIGURE: *Modèle graphique de l'IFA sans bruit appliquée aux CdVs avec contraintes spatiales.*



- cadre *semi-supervisé* : 1 label = 1 classe ou pas de classe

# DEUX A PRIORI

- a priori sur les dépendances entre variables observées et sources

- ▶ masque sur la matrice de mixage  $A$
- ▶ évite les permutations entre les sources

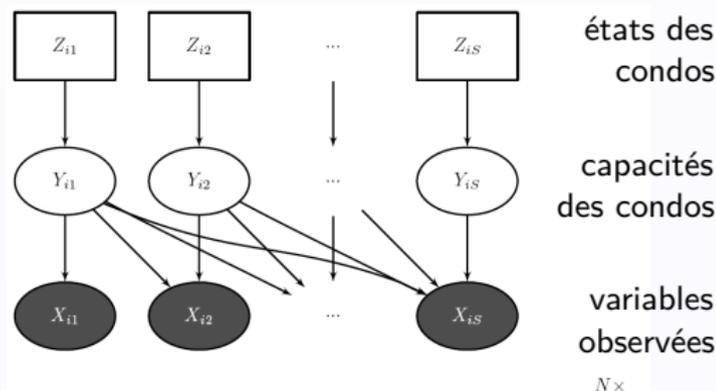


FIGURE: Modèle graphique de l'IFA sans bruit appliquée aux CdVs **avec** contraintes spatiales.

- cadre *semi-supervisé* : 1 label = 1 classe ou pas de classe

## Log-vraisemblance "exacte" des données observées

$$\mathcal{L}_1(\Phi; \mathbb{X}) = -N \cdot \log |\det(A)| + \sum_{i=1}^M \sum_{s=1}^S \sum_{k_s=1}^{K_s} \mathbb{1}_{Z_{is}=C_{k_s}} \cdot \log(\pi_{k_s} \cdot \mathcal{N}(y_{is}; \theta_{k_s}))$$

$$+ \sum_{i=M+1}^N \sum_{s=1}^S \log \left( \sum_{k_s=1}^{K_s} \pi_{k_s} \cdot \mathcal{N}(y_{is}; \theta_{k_s}) \right)$$

## EXEMPLE DÉTAILLÉ

- Soit  $x_i$ , un CdV composé de 3 condensateurs ( $L = 3/S = 6$ ).

$$x_i = y_i \cdot A^T$$



## EXEMPLE DÉTAILLÉ

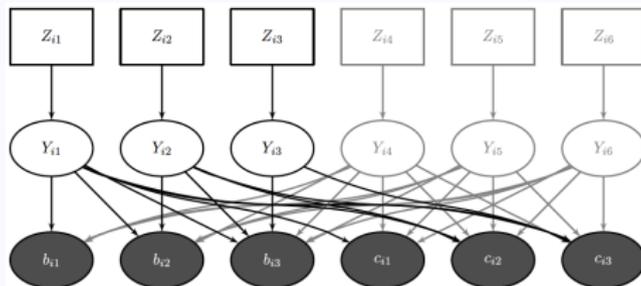
$$A^T = \begin{array}{|c|c|c|c|c|c|} \hline \star & \star & \star & \star & \star & \star \\ \hline 0 & \star & \star & 0 & \star & \star \\ \hline 0 & 0 & \star & 0 & 0 & \star \\ \hline \star & \star & \star & \star & \star & \star \\ \hline \star & \star & \star & \star & \star & \star \\ \hline \star & \star & \star & \star & \star & \star \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|c|c|c|} \hline y_{i1} & y_{i2} & y_{i3} & y_{i4} & y_{i5} & y_{i6} \\ \hline \end{array} \quad \begin{array}{|c|c|c|c|c|c|} \hline b_{i1} & b_{i2} & b_{i3} & c_{i1} & c_{i2} & c_{i3} \\ \hline \end{array}$$

$$\parallel$$
  

$$y_i$$

$$\parallel$$
  

$$x_i$$


$$L = 3/S = 6$$

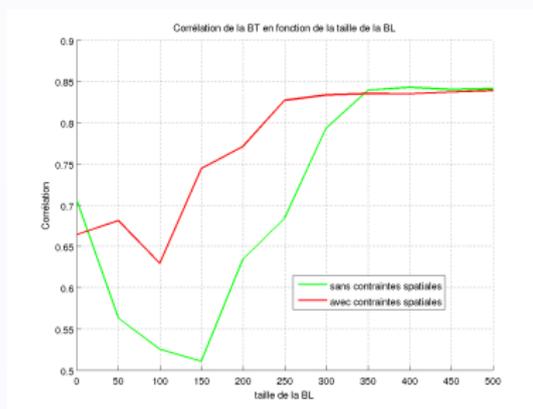
3 sources pertinentes

3 sources-bruits

$$\Phi = (A, (\pi_s, \theta_s)_{s \in \{1, \dots, 6\}})$$

# RÉSULTATS OBTENUS

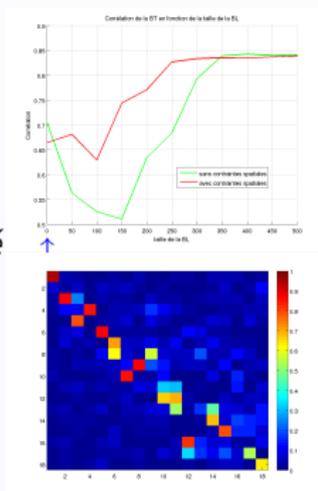
- Corrélation sur les sources :  
avec/sans contraintes spatiales



#condos = 18  
#BA = 500  
#BT = 2000

# RÉSULTATS OBTENUS

- Corrélation sur les sources :  
avec/sans contraintes spatiales  
 #BL=0 *non supervisé*



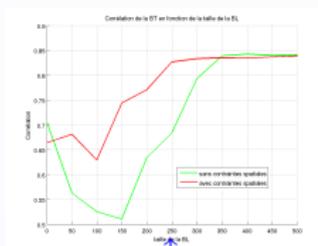
#condos = 18  
 #BA = 500  
 #BT = 2000

# RÉSULTATS OBTENUS

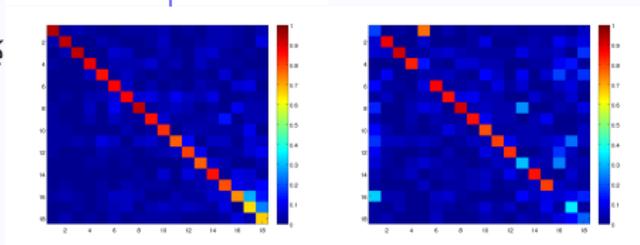
- Corrélation sur les sources :  
avec/sans contraintes spatiales

#BL=250

*semi-supervisé*

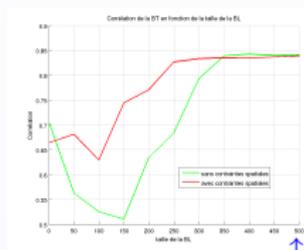


#condos = 18  
#BA = 500  
#BT = 2000



# RÉSULTATS OBTENUS

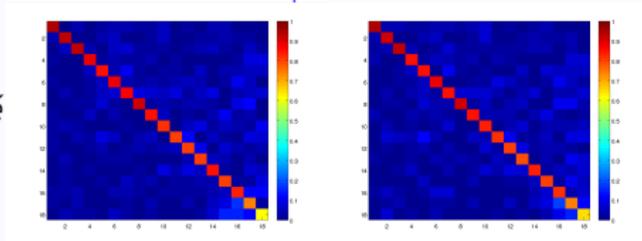
- Corrélation sur les sources :  
avec/sans contraintes spatiales



#condos = 18  
#BA = 500  
#BT = 2000

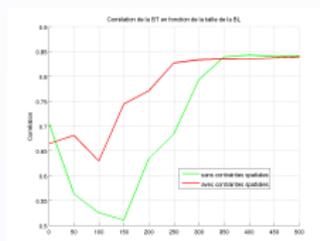
#BL=500

*supervisé*



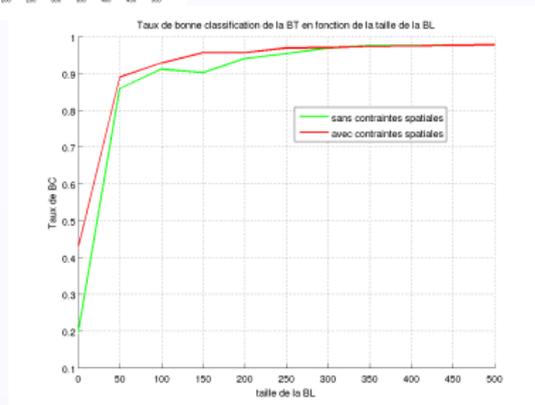
# RÉSULTATS OBTENUS

- Corrélation sur les sources :  
avec/sans contraintes spatiales



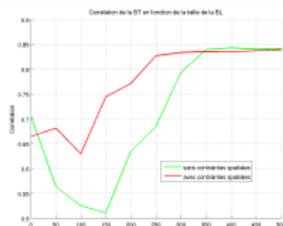
#condos = 18  
#BA = 500  
#BT = 2000

- Taux de bonne classification 1 :



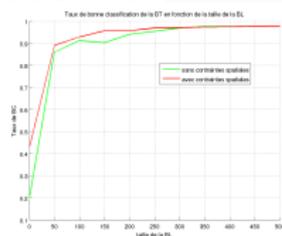
# RÉSULTATS OBTENUS

- Corrélation sur les sources :  
avec/sans contraintes spatiales

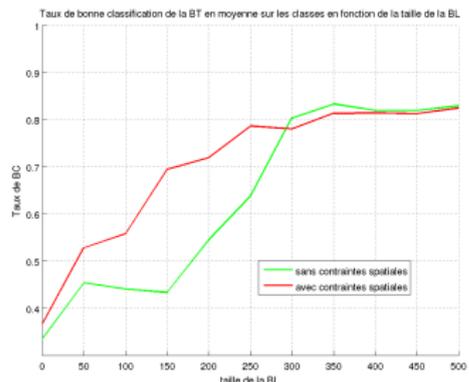


#condos = 18  
#BA = 500  
#BT = 2000

- Taux de bonne classification 1 :



- Taux de bonne classification 2 :



# EXTENSION À UN NOMBRE VARIABLE DE SOURCES (1)

- Apprentissage sur des CdVs de différentes tailles ( $L$  variable)

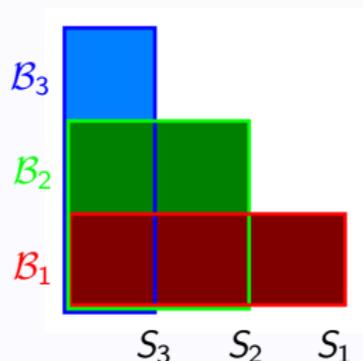
# EXTENSION À UN NOMBRE VARIABLE DE SOURCES (1)

- Apprentissage sur des CdVs de différentes tailles ( $L$  variable)
- “Trois bases d'apprentissage pour un seul apprentissage.”

$$\mathcal{B}_1 \text{ (} D_1 \times S_1 \text{)}$$

$$\mathcal{B}_2 \text{ (} D_2 \times S_2 \text{)}$$

$$\mathcal{B}_3 \text{ (} D_3 \times S_3 \text{)}$$



# EXTENSION À UN NOMBRE VARIABLE DE SOURCES (1)

- Apprentissage sur des CdVs de différentes tailles ( $L$  variable)
- “Trois bases d'apprentissage pour un seul apprentissage.”

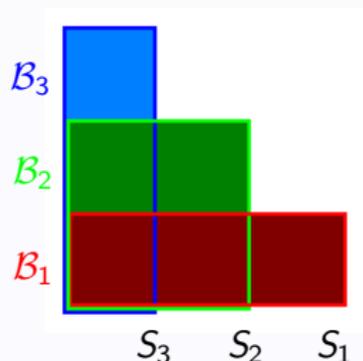
$$\mathcal{B}_1 \text{ (} D_1 \times S_1 \text{)}$$

$$\mathcal{B}_2 \text{ (} D_2 \times S_2 \text{)}$$

$$\mathcal{B}_3 \text{ (} D_3 \times S_3 \text{)}$$

Paramètre global

$$\Phi = (A_1, A_2, A_3, (\pi_s, \theta_s)_{s \in \{1, \dots, S_3, \dots, S_2, \dots, S_1\}})$$



# EXTENSION À UN NOMBRE VARIABLE DE SOURCES (1)

- Apprentissage sur des CdVs de différentes tailles ( $L$  variable)
- “Trois bases d'apprentissage pour un seul apprentissage.”

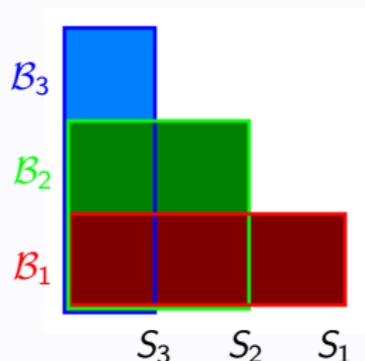
$$\mathcal{B}_1 \quad (D_1 \times S_1)$$

$$\mathcal{B}_2 \quad (D_2 \times S_2)$$

$$\mathcal{B}_3 \quad (D_3 \times S_3)$$

## Paramètre global

$$\Phi = (A_1, A_2, A_3, (\pi_s, \theta_s)_{s \in \{1, \dots, S_3, \dots, S_2, \dots, S_1\}})$$



- Construction de trois nouvelles bases d'apprentissage :

$$\mathbb{X}_1 \quad (N_1 \times S_1)$$

$$\mathbb{X}_2 \quad (N_2 \times S_2)$$

$$\mathbb{X}_3 \quad (N_3 \times S_3)$$

avec  $N_1 = D_1$ ,  $N_2 = D_1 + D_2$ , et  $N_3 = D_1 + D_2 + D_3$ .

# EXTENSION À UN NOMBRE VARIABLE DE SOURCES (2)

- Optimisation par un algorithme **GEM** :

- ▶ paramétrisation des sources

$$\underbrace{1, \dots, L_3}_{X_3}, \underbrace{L_3 + 1, \dots, L_2}_{X_2}, \underbrace{L_2 + 1, \dots, L_1}_{X_1}$$

- ▶ 3 montées de gradient pour les matrices de mixage  
↔ sauve les relations de dépendances/indépendances

# EXTENSION À UN NOMBRE VARIABLE DE SOURCES (2)

- Optimisation par un algorithme GEM :

- ▶ paramétrisation des sources

$$\underbrace{1, \dots, L_3}_{\mathbb{X}_3}, \underbrace{L_3 + 1, \dots, L_2}_{\mathbb{X}_2}, \underbrace{L_2 + 1, \dots, L_1}_{\mathbb{X}_1}$$

- ▶ 3 montées de gradient pour les matrices de mixage  
 ↪ sauve les relations de dépendances/indépendances

## Log-vraisemblance des données observées

$$\mathcal{L}_3(\Phi; \mathbb{X}_1, \dots, \mathbb{X}_V) =$$

$$\sum_{v=1}^V -N_v \cdot \log |\det(A_v)| + \sum_{i_v=1}^{M_v} \sum_{s=1}^{S_v} \sum_{k_s=1}^{K_s} \mathbb{1}_{Z_{i_v s} = C_{k_s}} \cdot \log (\pi_{k_s} \cdot \mathcal{N}(y_{i_v s}; \theta_{k_s}))$$

$$+ \sum_{i_v=M_v+1}^{N_v} \sum_{s=1}^{S_v} \log \left( \sum_{k_s=1}^{K_s} \pi_{k_s} \cdot \mathcal{N}(y_{i_v s}; \theta_{k_s}) \right)$$

# RÉSULTATS OBTENUS (1)

- 3 bases d'apprentissage avec des observations de différentes tailles :
  - ▶  $\mathcal{B}_1$  (500 × 48)
  - ▶  $\mathcal{B}_2$  (300 × 40)
  - ▶  $\mathcal{B}_3$  (300 × 36)

# RÉSULTATS OBTENUS (1)

- 3 bases d'apprentissage avec des observations de différentes tailles :

$$\begin{aligned} \blacktriangleright \mathcal{B}_1 \ (500 \times 48) &\Rightarrow \boxed{\mathbb{X}_1} \ (500 \times 48) \\ \blacktriangleright \mathcal{B}_2 \ (300 \times 40) &\Rightarrow \boxed{\mathbb{X}_2} \ (800 \times 40) \\ \blacktriangleright \mathcal{B}_3 \ (300 \times 36) &\Rightarrow \boxed{\mathbb{X}_3} \ (1100 \times 36) \end{aligned}$$

# RÉSULTATS OBTENUS (1)

- 3 bases d'apprentissage avec des observations de différentes tailles :

$$\begin{aligned} \blacktriangleright \mathcal{B}_1 \ (500 \times 48) &\Rightarrow \boxed{\mathbb{X}_1} \ (500 \times 48) \\ \blacktriangleright \mathcal{B}_2 \ (300 \times 40) &\Rightarrow \boxed{\mathbb{X}_2} \ (800 \times 40) \\ \blacktriangleright \mathcal{B}_3 \ (300 \times 36) &\Rightarrow \boxed{\mathbb{X}_3} \ (1100 \times 36) \end{aligned}$$

- # BT = 2000

# RÉSULTATS OBTENUS (1)

- 3 bases d'apprentissage avec des observations de différentes tailles :

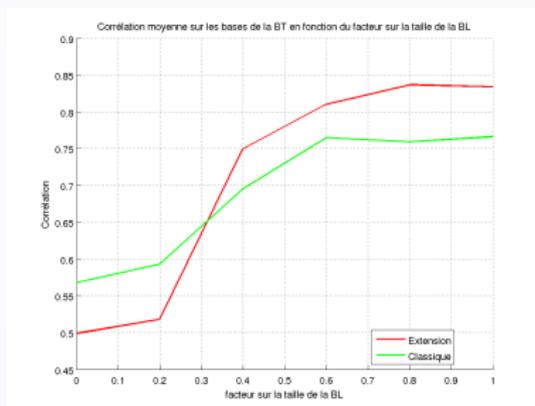
▶  $\mathcal{B}_1$  ( $500 \times 48$ )  $\Rightarrow$   $\mathbb{X}_1$  ( $500 \times 48$ )

▶  $\mathcal{B}_2$  ( $300 \times 40$ )  $\Rightarrow$   $\mathbb{X}_2$  ( $800 \times 40$ )

▶  $\mathcal{B}_3$  ( $300 \times 36$ )  $\Rightarrow$   $\mathbb{X}_3$  ( $1100 \times 36$ )

- # BT = 2000

- Corrélation sur les sources :  
extension/classique



# RÉSULTATS OBTENUS (1)

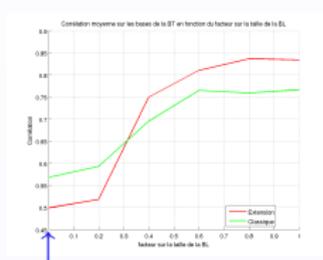
- 3 bases d'apprentissage avec des observations de différentes tailles :

▶  $B_1$  ( $500 \times 48$ )  $\Rightarrow$   $X_1$  ( $500 \times 48$ )

▶  $B_2$  ( $300 \times 40$ )  $\Rightarrow$   $X_2$  ( $800 \times 40$ )

▶  $B_3$  ( $300 \times 36$ )  $\Rightarrow$   $X_3$  ( $1100 \times 36$ )

- # BT = 2000

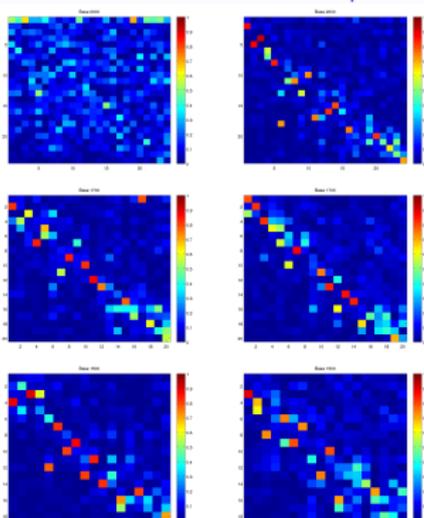


- Corrélation sur les sources :

extension/classique

#factBL=0

*non supervisé*



#condos = 24

#condos = 20

#condos = 18

# RÉSULTATS OBTENUS (1)

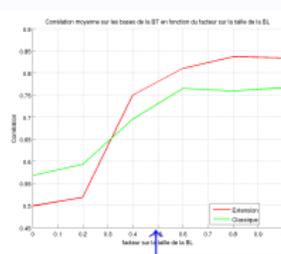
- 3 bases d'apprentissage avec des observations de différentes tailles :

▶  $B_1$  ( $500 \times 48$ )  $\Rightarrow$   $X_1$  ( $500 \times 48$ )

▶  $B_2$  ( $300 \times 40$ )  $\Rightarrow$   $X_2$  ( $800 \times 40$ )

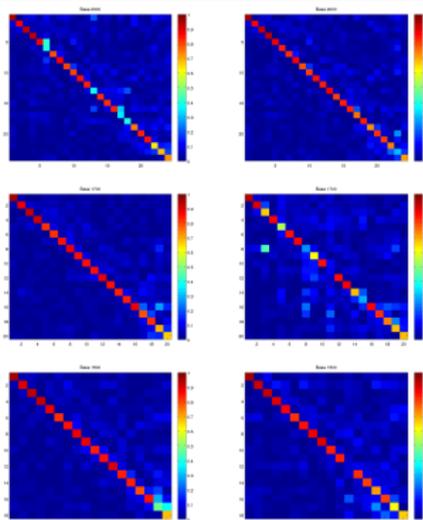
▶  $B_3$  ( $300 \times 36$ )  $\Rightarrow$   $X_3$  ( $1100 \times 36$ )

- # BT = 2000



- Corrélation sur les sources :  
extension/classique

#factBL=0.5    *semi-supervisé*



#condos = 24

#condos = 20

#condos = 18

# RÉSULTATS OBTENUS (1)

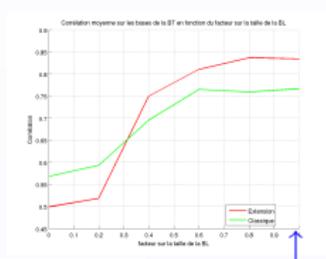
- 3 bases d'apprentissage avec des observations de différentes tailles :

▶  $B_1$  ( $500 \times 48$ )  $\Rightarrow$   $X_1$  ( $500 \times 48$ )

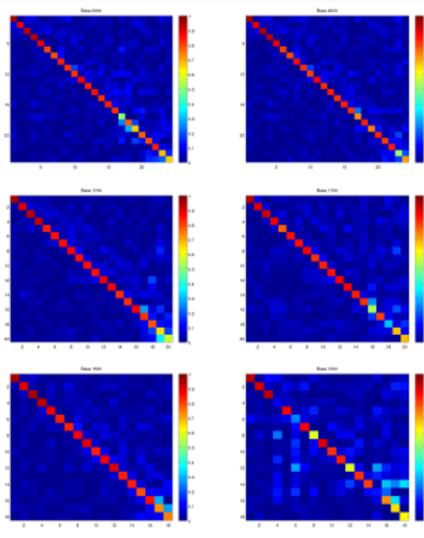
▶  $B_2$  ( $300 \times 40$ )  $\Rightarrow$   $X_2$  ( $800 \times 40$ )

▶  $B_3$  ( $300 \times 36$ )  $\Rightarrow$   $X_3$  ( $1100 \times 36$ )

- # BT = 2000



- Corrélation sur les sources :  
extension/classique



#condos = 24

#condos = 20

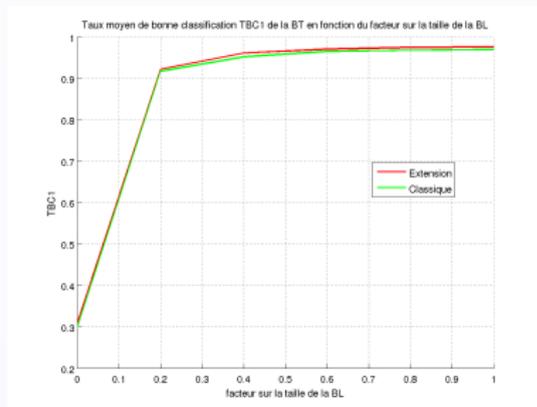
#condos = 18

#factBL=1

supervisé

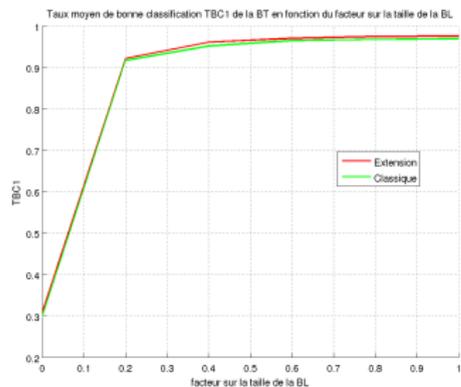
# RÉSULTATS OBTENUS (2)

- Taux de bonne classification 1 :

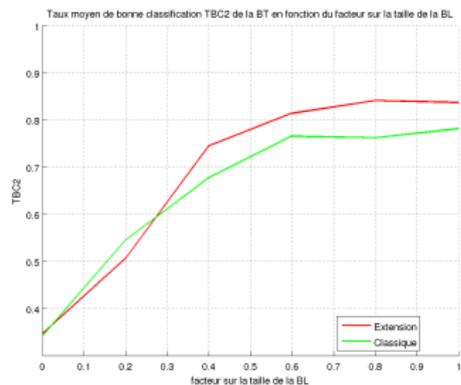


# RÉSULTATS OBTENUS (2)

- Taux de bonne classification 1 :



- Taux de bonne classification 2 :



# RAPPEL DE PLAN

- 1 Introduction
  - Problématique
  
- 2 Modèle de mélanges et Algorithme EM
  - Modèle de mélanges
  - Algorithme EM
  
- 3 Analyse en Facteurs Indépendants
  - Principe
  - Cadre du problème
  - Application : circuits de voie
  - Deux a priori
  - Extension à un nombre variable de sources
  
- 4 Conclusion

# CONCLUSION

- Synthèse :
  - ▶ **présentation théorique de l'IFA sans bruit, avec justification**
  - ▶ **application aux circuits de voie ferroviaires**
  - ▶ ajout de deux a priori au modèle (semi-supervisé & dépendances spatiales)
  - ▶ implémentation de plusieurs variantes du modèle
  - ▶ extension à un nombre variable de sources

# CONCLUSION

- Synthèse :
  - ▶ présentation théorique de l'IFA sans bruit, avec justification
  - ▶ application aux circuits de voie ferroviaires
  - ▶ ajout de deux a priori au modèle (semi-supervisé & dépendances spatiales)
  - ▶ implémentation de plusieurs variantes du modèle
  - ▶ extension à un nombre variable de sources

# CONCLUSION

- Synthèse :
  - ▶ présentation théorique de l'IFA sans bruit, avec justification
  - ▶ application aux circuits de voie ferroviaires
  - ▶ ajout de deux a priori au modèle (semi-supervisé & dépendances spatiales)
  - ▶ implémentation de plusieurs variantes du modèle
  - ▶ extension à un nombre variable de sources

# CONCLUSION

- Synthèse :
  - ▶ présentation théorique de l'IFA sans bruit, avec justification
  - ▶ application aux circuits de voie ferroviaires
  - ▶ ajout de deux a priori au modèle (semi-supervisé & dépendances spatiales)
  - ▶ implémentation de plusieurs variantes du modèle
  - ▶ extension à un nombre variable de sources
- Perspectives :
  - ▶ lenteur de la recherche linéaire  $\Rightarrow$  méthode de Quasi-Newton à mémoire limitée
  - ▶ relation non linéaire entre  $\mathbb{X}$  et  $\mathbb{Y}$
  - ▶ agréger d'autres critères à la vraisemblance (forme de  $p(\mathbb{Y}), \dots$ )

# BIBLIOGRAPHIE

-  E. Côme  
*Apprentissage de modèles génératifs pour le diagnostic de systèmes complexes avec labellisation douce et contraintes spatiales, PhDThesis, 2009*  
<http://www.inrets.fr/ur/ltn/poles/diag/pdf/TheseCome.pdf>
-  A. Dempster, N. Laird and D. Rubin  
*Maximum-likelihood from incomplete data via the em algorithm, 1977*  
<http://web.mit.edu/6.435/www/Dempster77.pdf>
-  H. Attias  
*Independent Factor Analysis, 1999*  
<http://research.goldenmetallic.com/ifa.pdf>
-  A. Hyvärinen and J. Karhunen and E. Oja  
*Independant Component Analysis, 2001*
-  G. McLachlan and T. Krishnan  
*The EM Algorithm and Extensions, 1997*
-  G. McLachlan and D. Peel  
*Finite Mixture Models, 2000*