
Approche bayésienne en séparation de sources. Applications en imagerie.

Hichem SNOUSSI

Groupe Problèmes Inverses

Laboratoire des Signaux et Systèmes
CNRS – SUPÉLEC – UPS

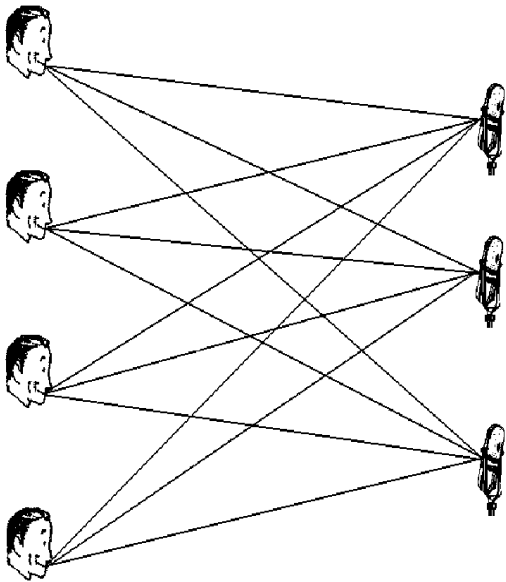
Directeur de thèse : Ali Mohammad-Djafari

Gif-sur-Yvette, 29 septembre 2003

Plan de la présentation

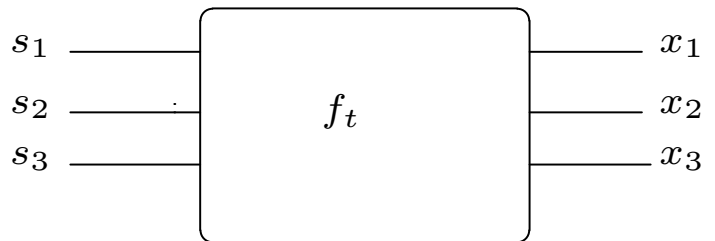
- ❑ Position du problème.
- ❑ Méthodologie bayésienne / Théorie de l'information
- ❑ Approche bayésienne.
 - Modélisation des sources.
 - Dégénérescence du maximum de vraisemblance.
 - Géométrie de l'information et sélection d'*a priori*.
- ❑ Algorithmes de séparation : mise en œuvre et applications.
 - Exploitation de la non stationnarité temporelle, spatiale, spectrale...
 - Application en imagerie satellitaire.
 - Application en cosmologie observationnelle.
- ❑ Conclusions et perspectives

Position du problème



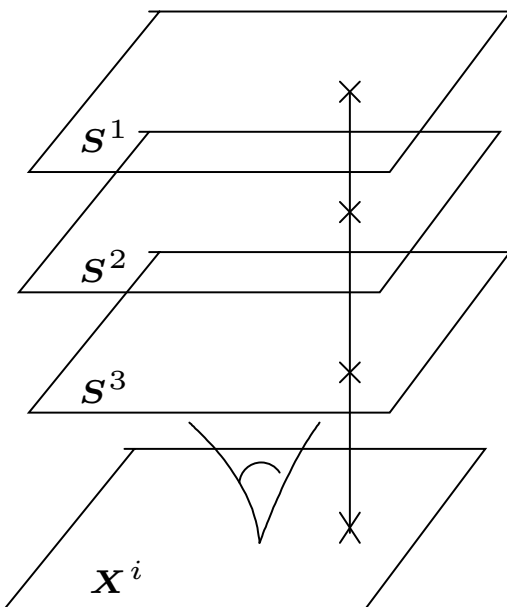
Mélange de sons

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



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

Mélange d'images



Mélange de composantes astrophysiques

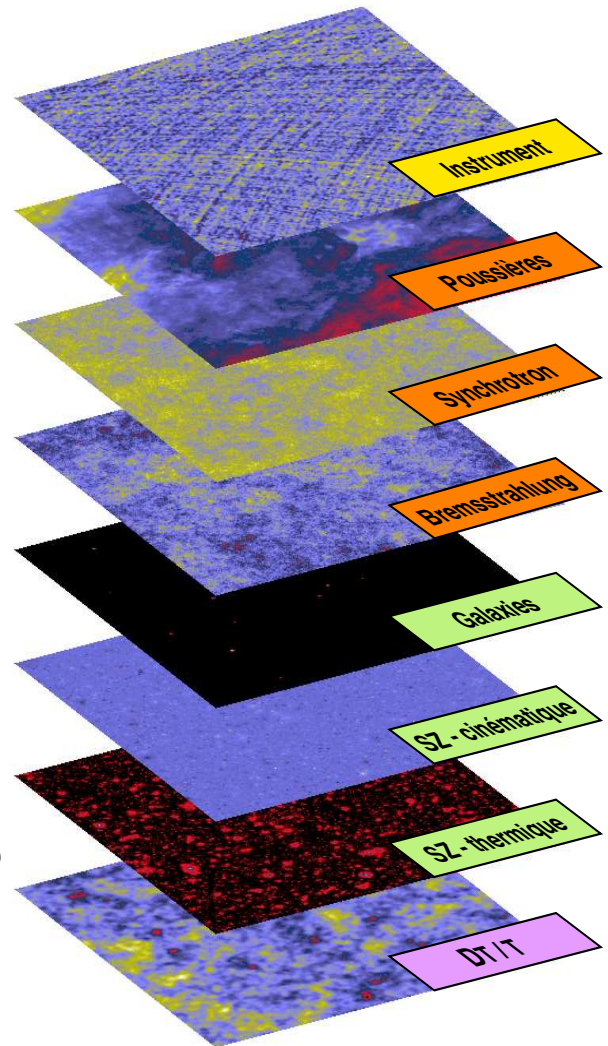
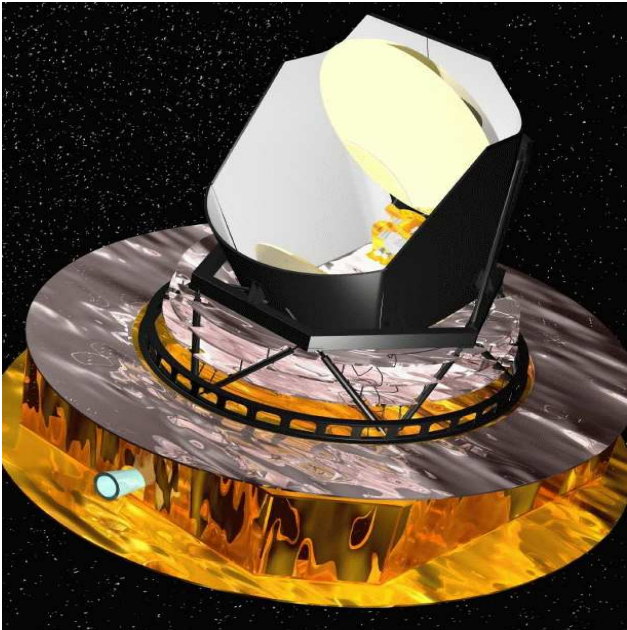
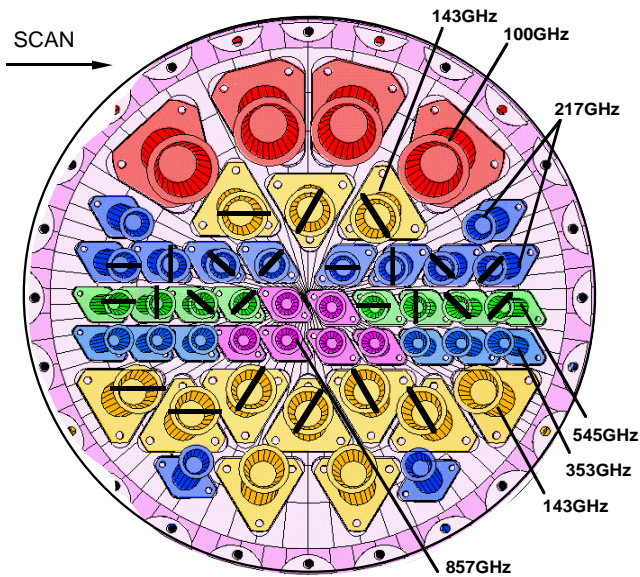
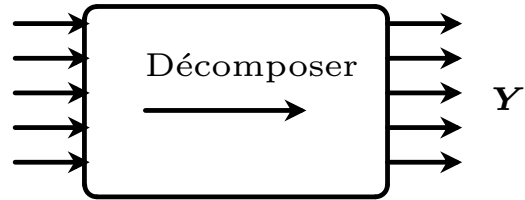
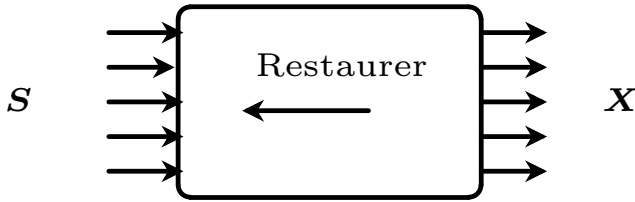


figure RUMBA, 1996



$$x = As + b.$$

Objectifs et méthodologies



Méthodologie bayésienne

Théorie de l'information

$$x = As + b$$

$$A, s_{1..T}$$

$$A$$

$$s_{1..T}$$

$$\theta$$

$$\mathcal{I} := (x_{1..T}, I \longrightarrow)$$



Introduction des probabilités

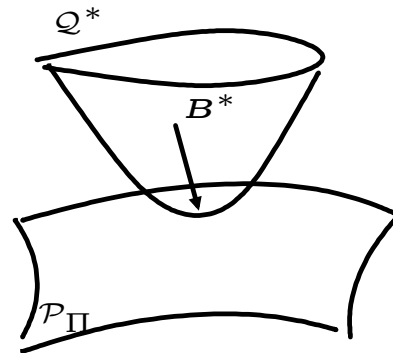


$$Pr(x_{1..T}, I \longrightarrow A, s_{1..T}, \theta)$$

distribution a posteriori

→ consistance du calcul des probabilités

$y = Bx, x = As$
 y à composantes indépendantes
 (ACI)



Mesure de l'indépendance par
 la divergence de Kullback-Leibler

$$KL(p(y), \prod p_j(y_j))$$

Point de rencontre : $b \rightarrow 0$

- MV \implies Information mutuelle
- Théorème de Darmois $\longrightarrow y = Bx = BA s$

Approche bayésienne

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

- **Distribution *a posteriori*** (règle de Bayes) :
[Mohammad-Djafari99, Knuth99]

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

- Structure naturelle à **variables cachées** :

$\mathbf{x}_{1..T}$ \longrightarrow les données incomplètes.

$\mathbf{s}_{1..T}$ \longrightarrow les données manquantes.

- Relation avec l'analyse en composantes indépendantes (ACI) :

- Prise en compte du **bruit** dans le modèle.
- Prise en compte des informations *a priori* sur les coefficients du mélange \longrightarrow **régularisation** de l'ACI dans le cas non bruité.

- **Choix des probabilités** pour le bruit, les sources, la matrice de mélange... ?

Modélisation des sources

Modélisation des sources

- i.i.d non gaussiennes [Gaeta90, Jutten91, Bermond00]
- Gaussiennes corrélées [Belouchrani95]
- Gaussiennes non stationnaires [Pham01]
- [Snoussi01d] : **Mélange de distributions**

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

□ Processus **doublement** stochastique :

Sources réelles	\mathbf{s}_1	\mathbf{s}_2	\mathbf{s}_3	\dots	\mathbf{s}_T
	↑	↑	↑	\dots	↑
Sources cachées	\mathbf{z}_1	\mathbf{z}_2	\mathbf{z}_3	\dots	\mathbf{z}_T

- Connaissant les variables $\mathbf{z}_{1..T}$, les sources sont temporellement indépendantes :

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

- Les \mathbf{z}_t prennent des valeurs discrètes (classes) et sont :

→ soit **indépendantes** (mélange i.i.d),

→ soit **corrélées** avec une structure **Markovienne** (chaîne de Markov 1-D, champ de Markov 2-D).

Modélisation des sources

$$p(\mathbf{s}_{1..T} | \boldsymbol{\theta}) = \sum_{z_{1..T}} p(z_{1..T} | \boldsymbol{\pi}) \prod_{t=1}^T \mathcal{N}(\mathbf{s}_t; \boldsymbol{\mu}_{z_t}, \mathbf{R}_{z_t})$$

□ Avantages :

- Interprétation **Logique** : $\mathbf{I} \longrightarrow \mathbf{s}_{1..T} \rightsquigarrow \mathbf{I} \longrightarrow z_{1..T} \longrightarrow \mathbf{s}_{1..T} \longrightarrow$ cadre logique à l'introduction de la classification.
- Modélisation de la **dépendance** temporelle, spatiale, spectrale,...
- Même **structure cachée** que le problème de séparation de sources.
- Interprétation du critère \longrightarrow **exploitation de la non stationnarité** : classification + ajustement de matrices de covariance.
- Bonne alternative à la modélisation **non paramétrique**.

Dégénérescence
du maximum de vraisemblance

Dégénérescence du maximum de vraisemblance

☐ Sources **directement** observées :

$$p(\mathbf{s}_{1..T} | \boldsymbol{\theta}) = \sum_{z_{1..T}} p(z_{1..T} | \boldsymbol{\pi}) \prod_{t=1}^T \mathcal{N}(\mathbf{s}_t; \boldsymbol{\mu}_{z_t}, \mathbf{R}_{z_t})$$

↓

Maximum de vraisemblance

↓

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta} \{p(\mathbf{s}_{1..T} | \boldsymbol{\theta})\}.$$

→ **La vraisemblance n'est pas bornée.**

- Cas 1-D, étude de la dégénérescence ($\sigma_z \rightarrow 0$) + solution bayésienne (*a priori* inverse gamma) [Ridolfi00].
- Dans [Snoussi01c] :
 1. Caractérisation de l'ensemble des points de singularités dans le cas **vectorel**.
 2. Solution **bayésienne** → pénalisation avec un **inverse wishart** sur les \mathbf{R}_z .
 3. Tenir compte des contraintes de structure des matrices de covariance → **"Reversed EM"**

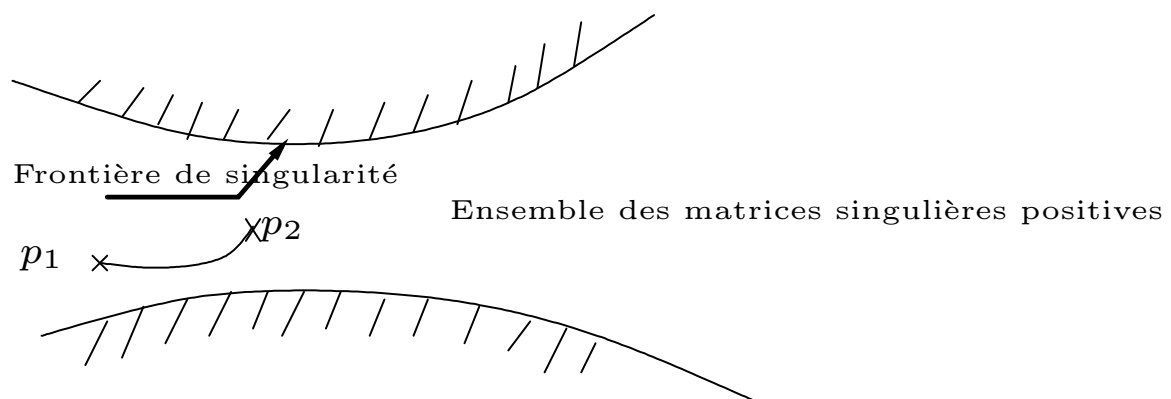
4. Problème de **dégénérescence** en séparation de sources :

$$\begin{cases} \mathbf{x}_{1..T} = \mathbf{A}\mathbf{s}_{1..T} + \mathbf{b}_{1..T} \\ p(\mathbf{s}_{1..T} | \boldsymbol{\theta}) \end{cases} \longrightarrow \text{Estimation de } \mathbf{A}, \mathbf{R}_\epsilon, \boldsymbol{\theta}$$

Condition nécessaire : \mathbf{R}_ϵ soit singulière + au moins une des \mathbf{R}_z soit singulière.

5. Pénalisation avec un inverse wishart sur la covariance du bruit élimine la dégénérescence.

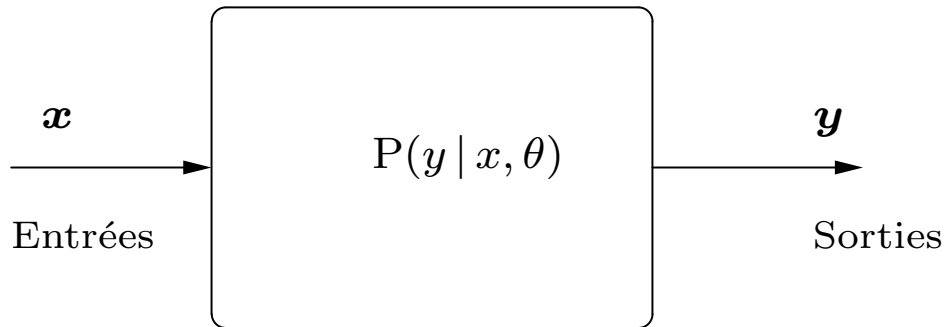
□ **Interprétation géométrique** : l'information de Fisher est singulière sur la frontière \longrightarrow **non différentiabilité** de la variété riemannienne.



Vers la Géométrie de l'information...

Géométrie de l'information et sélection d'*a priori*

Prédiction bayésienne



Système paramétrique $P(y | x, \theta)$

❑ Objectif

$\mathbf{z} = (x_i, y_i)$: données d'apprentissage \longrightarrow Prédiction des sorties y

Fonction de prédiction :

$$\tau : \mathbf{z} \longrightarrow \tau(\mathbf{z}) = q \in \mathcal{Q} \subset \mathcal{P} \subset \tilde{\mathcal{P}}$$

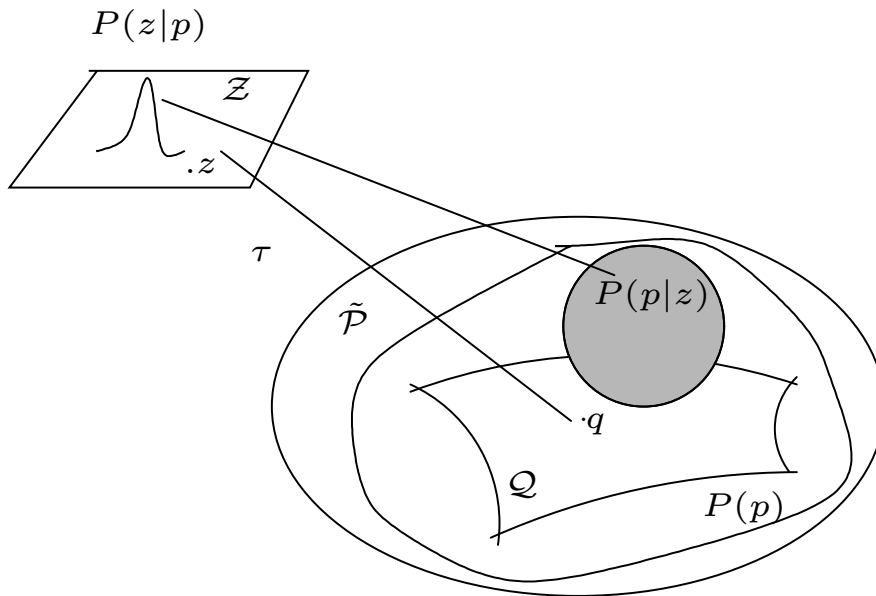
$\mathcal{Q} = \{P(z | \theta), \theta \in \Theta\} \longrightarrow$ Modèle paramétrique.

\mathcal{P} ensemble des mesures de probabilités.

$\tilde{\mathcal{P}}$ ensemble des mesures positives de masses finies.

Masse et géométrie

□ Masse = distribution *a posteriori*



• Règle de Bayes $\longrightarrow P(p | z) \propto P(z | p) P(p)$

□ Géométrie :

δ -divergence D_δ :

$$D_\delta(p, q) = \frac{\int p}{1 - \delta} + \frac{\int q}{\delta} - \frac{\int p^\delta q^{1-\delta}}{\delta(1 - \delta)}$$

D_δ induit une structure **duale** $(g, \nabla^\delta, \nabla^{1-\delta})$ [Amari85]

$g \longrightarrow$ métrique de Fisher

$\nabla^\delta \longrightarrow \delta$ connection.

Prédiction bayésienne

❑ Erreur de généralisation

$$E_\delta(\tau) = \int_p P(p) \int_z P(z | p) D_\delta(p, \tau(z))$$

- Fonction optimale $\tau_\delta = \arg \min_\tau E_\delta(\tau)$

❑ Cohérence de l'estimateur optimale [Zhu95]

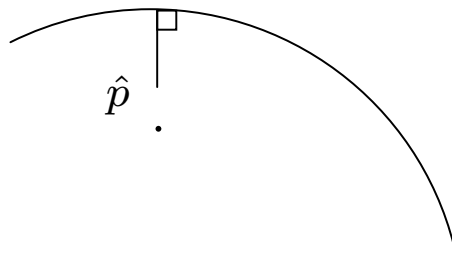
$$\hat{p} = \tau_\delta(z) = \arg \min_q \int_p P(p | z) D_\delta(p, q)$$

❑ L'estimateur optimale est le centre de gravité de \mathcal{P} muni de $(P(p | z), D_\delta)$:

$$\hat{p}^\delta \propto \int p^\delta P(p | z)$$

$\tilde{\mathcal{P}}$ est convexe $\forall \delta \in [0, 1]$

\mathcal{P} non convexe pour $\delta \in [0, 1[\longrightarrow$ projection de l'estimateur optimale :



Sélection d'*a priori*

❑ Résoudre le **problème variationnel** : [Snoussi02c]

$$J(\Pi) = \gamma_e \int \Pi(\theta) D_\delta(p_\theta, p_0) d\theta + \gamma_u \int \Pi(\theta) \log \Pi(\theta) / \sqrt{g(\theta)} d\theta$$

\downarrow
 δ -Divergence

\downarrow
 1-Divergence

$$J(\Pi) = E(\tau_0) + \gamma_u \int \Pi(\theta) \log \Pi(\theta) / \sqrt{g(\theta)} d\theta, \quad \frac{\partial \tau_0}{\partial z} = 0$$

$$\implies \boxed{\Pi_\delta(\theta) \propto \exp \left[-\frac{\gamma_e}{\gamma_u} D_\delta(p_\theta, p_0) \right] \sqrt{g(\theta)}}$$

❑ **Invariance** par rapport à la **paramétrisation** + **invariance** par rapport au choix de la **mesure dominante** dans l'espace des données.

- $\gamma_e = 0 \longrightarrow a \text{ priori}$ de Jeffreys.
- $\gamma_u = 0 \longrightarrow \Pi(\theta) = \delta_{p_0}$.
- $\delta = 1 \longrightarrow a \text{ priori}$ entropique [Rodriguez90].
- $\delta = 0 \longrightarrow a \text{ priori}$ conjugué pour la famille exponentielle.

familles δ -plates

- Définition [Amari85] :

\mathcal{Q} est une famille δ -plate \iff l'existence de coordonnées affines $[\theta]$ telle que la δ connection $\Gamma_\delta(\theta) = 0$.

- Propriété (structure duale) :

\mathcal{Q} est δ -plate $\iff \mathcal{Q}$ est $(1 - \delta)$ -plate.

- Propriété :

Les coordonnées δ -affines $[\theta]$ et $(1 - \delta)$ -affines $[\eta]$ sont liées par la transformation de Legendre et la divergence canonique s'écrit : is :

$$D_\delta(p, q) = \psi(p) + \phi(q) - \theta_i(p) \eta_i(q)$$

- Exemples :

La famille exponentielle est 1-plate.

Les mélanges de densités est une famille 0-plate.

$\tilde{\mathcal{P}} = \{p, \int p < \infty\}$ est δ -plat $\forall \delta \in [0, 1]$.

Sélection d'*a priori* pour les familles δ -plates

□ Expression analytique :

$$\implies \Pi(\theta) \propto \exp \left[-\frac{\gamma_e}{\gamma_u} (\psi(\theta) - \theta_i \eta_i^0) \right] \sqrt{g(\theta)}$$

□ Cas **euclidien** : $[\theta_i] = [\eta_i]$

$$\implies \Pi(\theta) \propto \exp \left[-\frac{\gamma_e}{\gamma_u} \|\theta - \theta_0\|^2 \right]$$

\implies **Gaussienne** de moyenne θ_0 et de précision $\frac{\gamma_e}{\gamma_u}$

□ Extension au cas de **mélange** de familles δ -plates : [Snoussi02c]

- Mélange de **gaussiennes vectorielles** :

1. Singularité \leftrightarrow non différentiabilité de la transformation de Legendre.

2. Comportement de $D_\delta(p_\theta, p_0)$ différent selon δ :

$\longrightarrow \delta = 0 : \Pi_0 \implies$ **inverse wishart**.

$\longrightarrow \delta = 1 : \Pi_1 \implies$ **wishart**.

□ Modèle à données manquantes : **séparation de sources**.

Algorithmes de séparation de sources : Exploitation de la non stationnarité

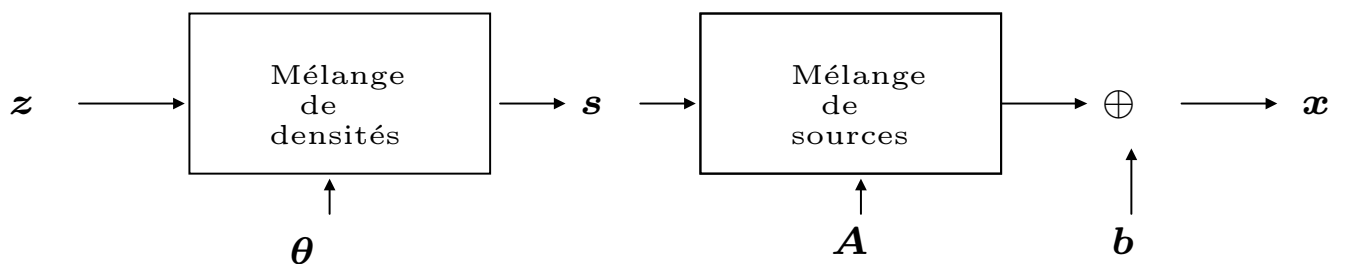
Structure doublement cachée

❑ Les sources (s_1, \dots, s_T) ne sont pas **directement** observées :

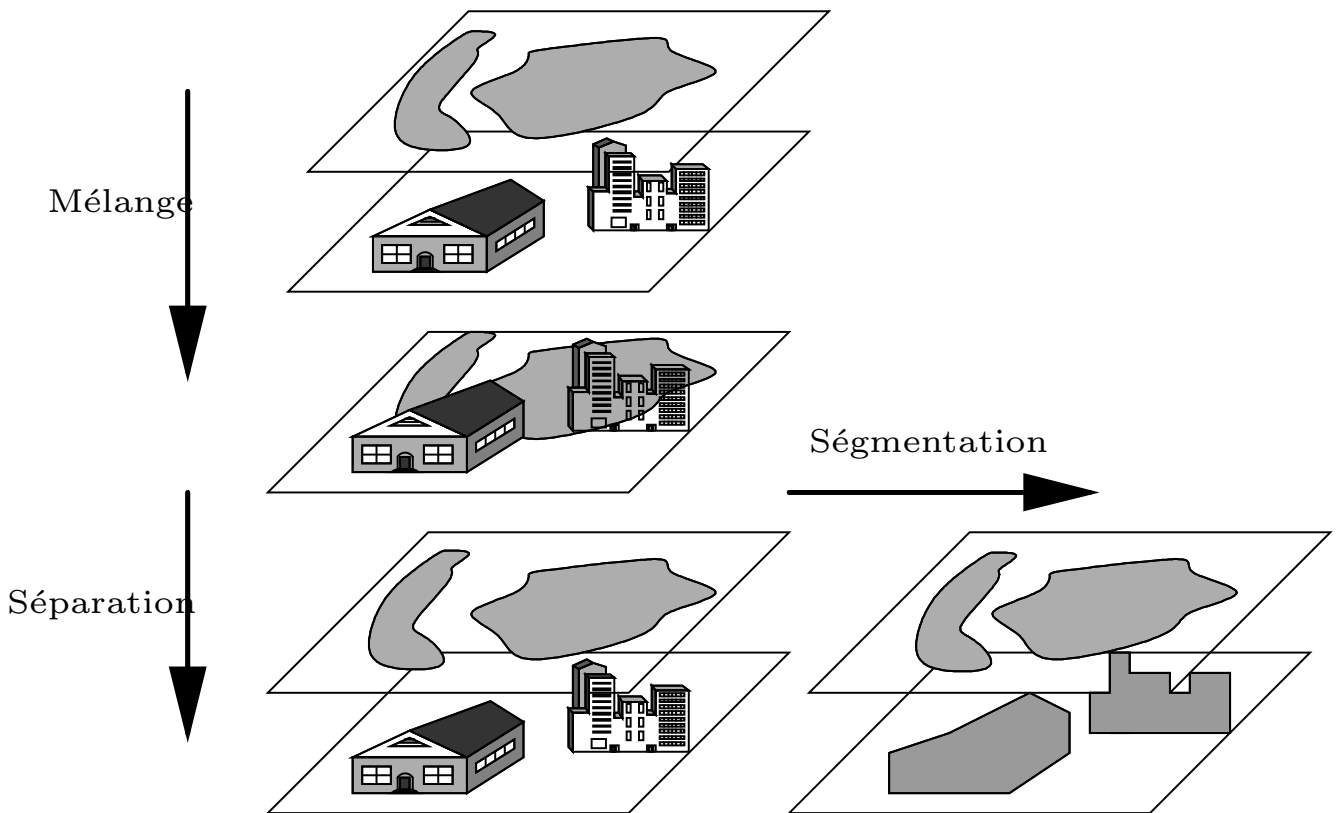
$$\left\{ \begin{array}{l} \mathbf{x}_{1..T} = \mathbf{A} \mathbf{s}_{1..T} + \mathbf{b}_{1..T}, \quad t = 1..T, \\ p(\mathbf{s}_{1..T}) = \sum_{\mathbf{z}_{1..T}} p(\mathbf{z}_{1..T})p(\mathbf{s}_{1..T} | \mathbf{z}_{1..T}) \end{array} \right.$$

→ (s_1, \dots, s_T) forment une deuxième couche de variables **cachées**.

Sources mélangées	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	...	\mathbf{x}_T
	↑	↑	↑	...	↑
Sources réelles	\mathbf{s}_1	\mathbf{s}_2	\mathbf{s}_3	...	\mathbf{s}_T
	↑	↑	↑	...	↑
Étiquettes cachées	z_1	z_2	z_3	...	z_T



Non stationnarité spatiale



☐ Forte **similarité** entre la séparation de sources et la ségmentation d'images : problèmes à variables cachées → Efficacité algorithmique.

→ Séparation et ségmentation **simultanées**.

☐ Interprétation du critère (connaissant $z_{1..T}$) :

$$\begin{aligned}
 \mathcal{J} &= \log p(\mathbf{A} | \mathbf{x}_{1..T}, \mathbf{z}_{1..T}) \\
 &= \underbrace{\sum_{z=1}^K \alpha_z D_{\mathcal{L}\mathcal{K}}(\mathbf{R}_{xx} || \mathbf{A}\mathbf{R}_z\mathbf{A}^T + \mathbf{R}_\epsilon)}_{\text{Ajustement de covariances [Pham01]}} + \underbrace{\log p(\mathbf{A})}_{\text{régularisation}}
 \end{aligned}$$

Aspects algorithmiques

$$\eta = (A, R_\epsilon, \theta)$$

☐ Algorithme EM → à partir de $\eta^{(0)}$, itérer :

(i) **E. (Expectation)** → calcul de la fonctionnelle :

$$Q(\eta | \eta^{(k-1)}) = E_{s,z} [\log p(\mathbf{x}, s, z | \eta) + \log p(\eta | \mathbf{x}, \eta^{(k-1)})]$$

(ii) **M. (Maximization)** → maximisation de la fonctionnelle :

$$\eta^{(k)} = \arg \max_{\eta} \left\{ Q(\eta | \eta^{(k-1)}) \right\}$$

Calcul des probabilités marginales $p(z_t | \mathbf{x}_{1..T}, \eta^{(k-1)})$

☐ Cas 1-D : [Snoussi02a]

- EM exact → procédure de Baum Welsh [$O([\prod_{j=1}^n K_j]^2 T)$].
- Approximations de l'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
réduction du coût dû à la structure markovienne		réduction du coût dû à la structure spatiale inter-capteurs

Aspects algorithmiques : cas 2-D

□ Approximations **stochastiques** de l'EM : [Snoussi02b]

1. Simuler M échantillons $\mathbf{Z}^{(m)}$ (M images \mathbf{Z}) selon $p(\mathbf{Z} | \mathbf{X}, \boldsymbol{\eta}^{(k-1)})$

2. Construire la fonctionnelle :

$$\tilde{Q}(\boldsymbol{\eta} | \boldsymbol{\eta}^{(k-1)}) = \frac{1}{M} \mathbb{E}_s \left[\log p(\mathbf{X}, \mathbf{S}, \mathbf{Z}^{(m)} | \boldsymbol{\eta}) \right] + \log p(\boldsymbol{\eta})$$

→ **Somme empirique** sur \mathbf{Z} et une **intégration exacte** par rapport à \mathbf{S} .

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

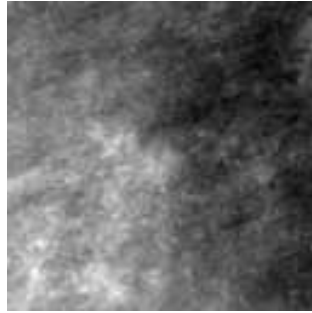


□ **Echantillonnage de Gibbs** :

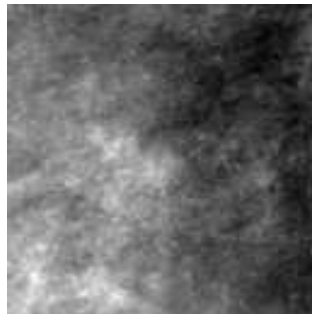
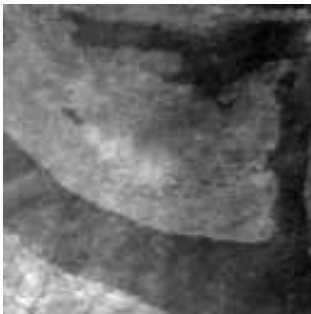
1. Simule $(\tilde{\mathbf{Z}}^{(k)}, \tilde{\mathbf{S}}^{(k)}) \sim p(\mathbf{Z}, \mathbf{S} | \mathbf{X}, \tilde{\boldsymbol{\eta}}^{(k-1)})$

2. Simule $\tilde{\boldsymbol{\eta}}^{(k)} \sim p(\boldsymbol{\eta} | \mathbf{X}, \tilde{\mathbf{Z}}^{(k)}, \tilde{\mathbf{S}}^{(k)})$

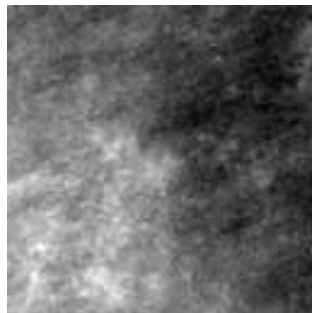
Séparation et ségmentation conjointe par MCMC



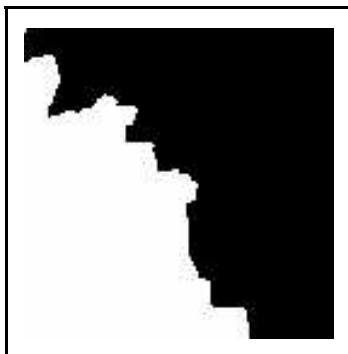
Sources originales



Sources mélangées



Sources séparées

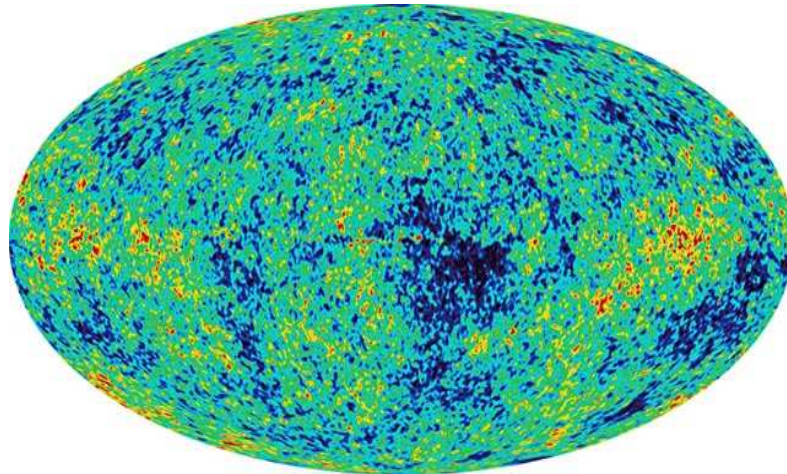


Sources ségmentées

Application en cosmologie observationnelle.

Collaboration avec l'IN2P3, Collège de France.

Observation du CMB



□ Le CMB représente les fluctuations statistiques sur le ciel (anisotropies) :

$$\frac{\Delta T}{T} \sim 10^{-5}$$

□ Objectif : Mesurer les anisotropies du CMB qui nous renseignent sur :

- Géométrie de l'univers.
- Formation des structures (dans l'univers).
- Paramètres cosmologiques (masse et âge de l'univers,...)

Mélange de composantes astrophysiques

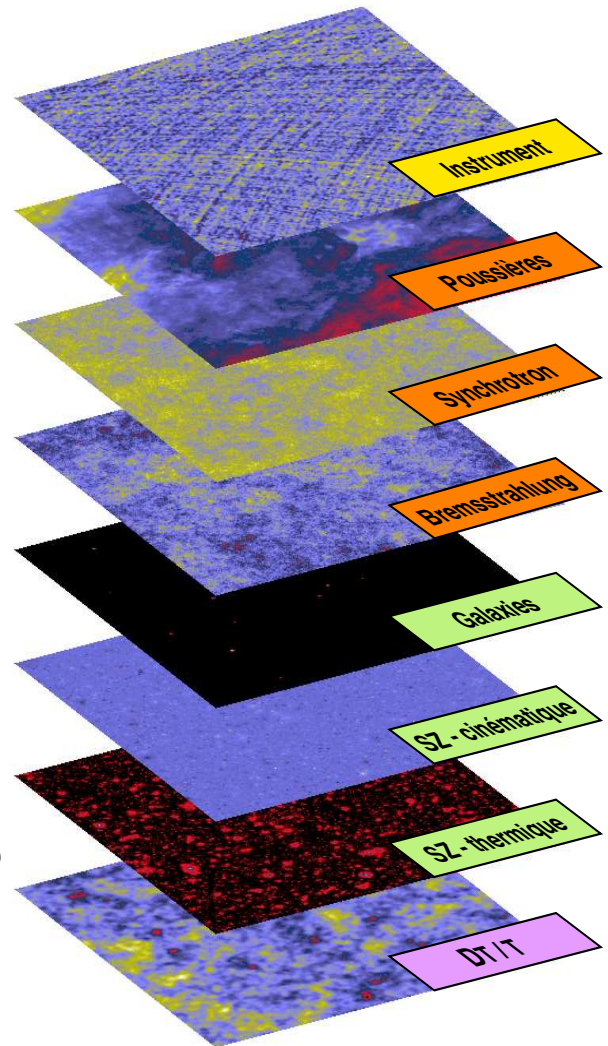
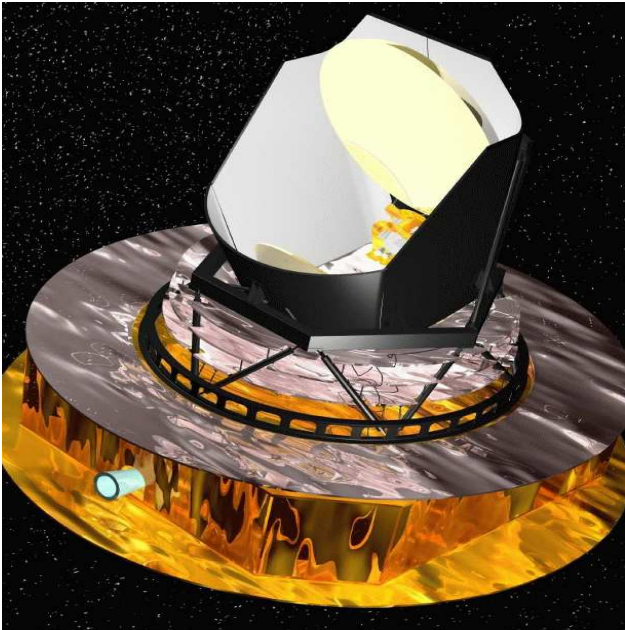
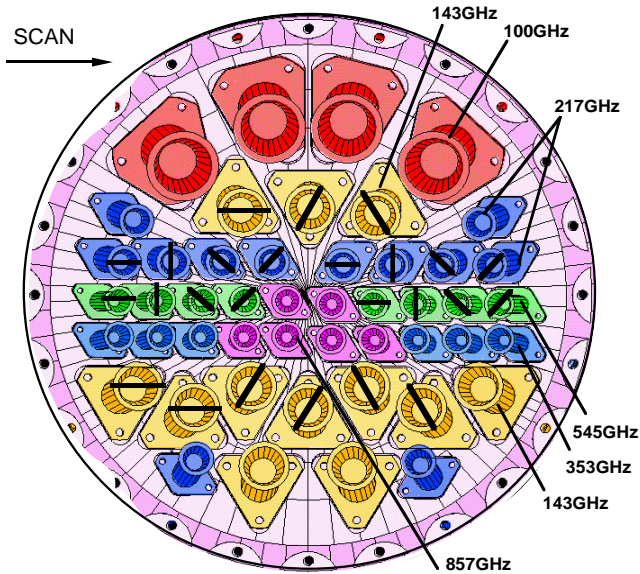


figure RUMBA, 1996



$$x = As + b.$$

Modélisation du mélange

Hypothèse : Les spectres électromagnétiques des composantes sont indépendants de la position sur le ciel :

$$F(\nu, \vec{r}) = g(\nu) \times s(\vec{r})$$

ν \longrightarrow fréquence électromagnétique.

Sur le détecteur d :

$$X_d(\vec{r}) = \sum_{i=1}^{N_c} \mathbf{A}_d^i \cdot s_i(\vec{r}) + n_d(\vec{r})$$

En pratique, les signaux captés sont convolués :

$$X_d(\vec{r}) = \sum_{i=1}^{N_c} \mathbf{A}_d^i \cdot \int s_i(\vec{r} - \vec{r}') \mathbf{B}_d(\vec{r}') d\vec{r}' + n_d(\vec{r})$$

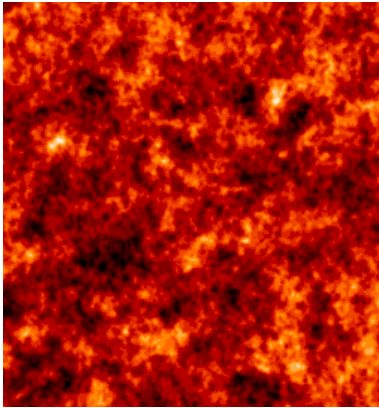
Dans l'espace de Fourier :

$$X_d(\vec{k}) = \sum_{i=1}^{N_c} \mathbf{A}_d^i \cdot \mathbf{B}_d(\vec{k}) \cdot s_i(\vec{k}) + n_d(\vec{k})$$

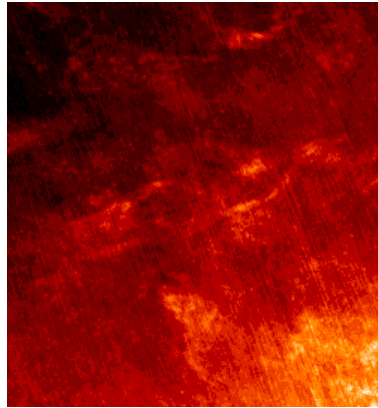
$$\mathbf{x}(\vec{k}) = \mathbf{A} \mathbf{s}(\vec{k}) + \mathbf{b}(\vec{k})$$

- Les trois composantes originales simulées :

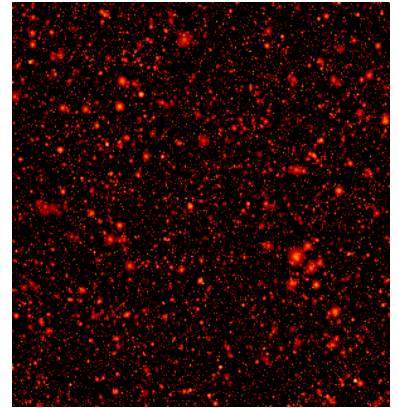
CMB



Dust

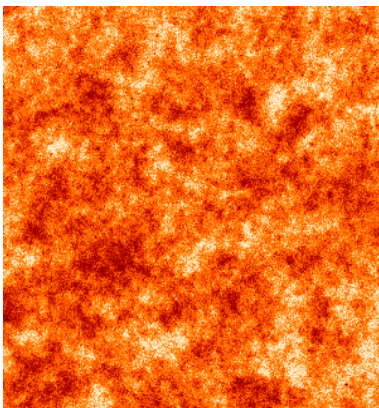


SZ

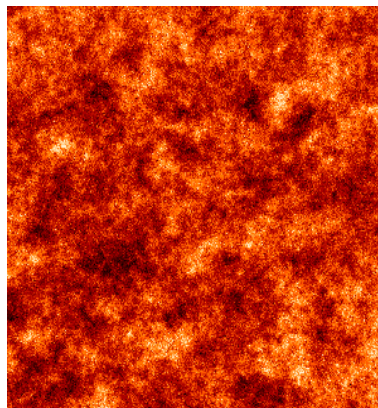


- Les composantes mélangées :

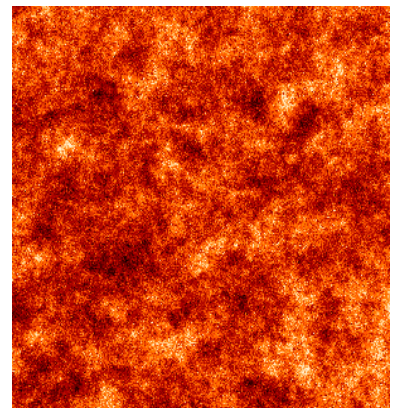
100 GHz



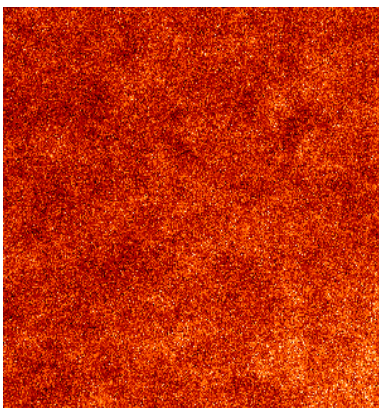
150 GHz



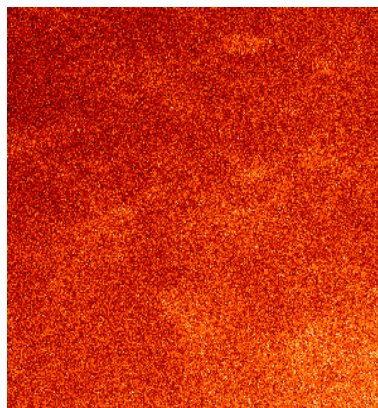
217 GHz



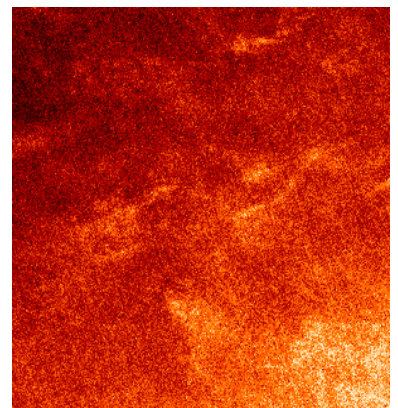
343 GHz



545 GHz



857 GHz



Modèle et hypothèses

$$\mathbf{x}(\vec{k}) = \mathbf{A} \mathbf{s}(\vec{k}) + \mathbf{b}(\vec{k})$$

□ Hypothèses :

- **Stationnarité** \implies **Décorrélation** dans le domaine spectral.

- **Gaussiannité** :

$p(\mathbf{b}_k) = \mathcal{N}(0, \mathbf{R}_\epsilon) \longrightarrow \mathbf{R}_\epsilon$ constante, diagonale et inconnue :

$p(\mathbf{s}_k) = \mathcal{N}(0, \mathbf{C}_k) \longrightarrow \mathbf{C}_k$ diagonale inconnue :

$$\left\{ \begin{array}{l} p(\mathbf{b}_{1..K}) = \prod_{k=1}^K p(\mathbf{b}_k) \\ p(\mathbf{s}_{1..K}) = \prod_{k=1}^K p(\mathbf{s}_k) \end{array} \right.$$

□ Problème à données **incomplètes** :

- $(\mathbf{x}_k)_{k=1..K} \longrightarrow$ données observées **incomplètes**.

- $(\mathbf{s}_k)_{k=1..K} \longrightarrow$ données **manquantes**.

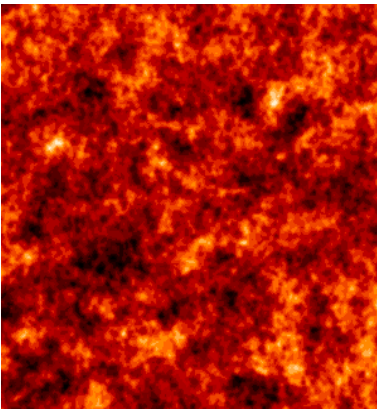
\longrightarrow Vraisemblance "complétée" $p(\mathbf{x}_{1..K}, \mathbf{s}_{1..K} \mid \boldsymbol{\theta})$.

- **Algorithme EM dans le domaine spectral** : [Snoussi01e].

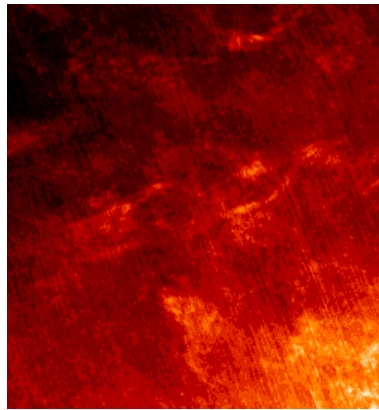
Résultats de séparation

Reconstruction des sources

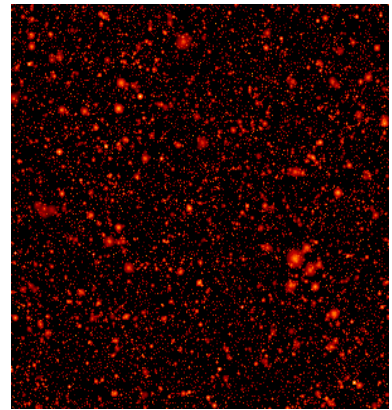
CMB



Dust



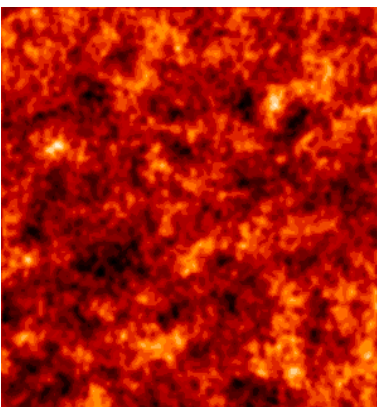
SZ



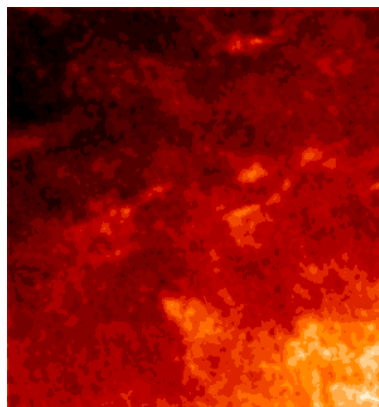
Reconstruction avec EM spectral



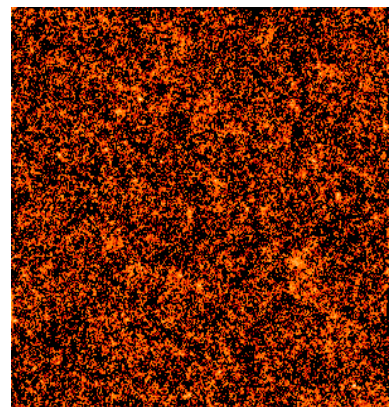
recovered CMB



recovered Dust

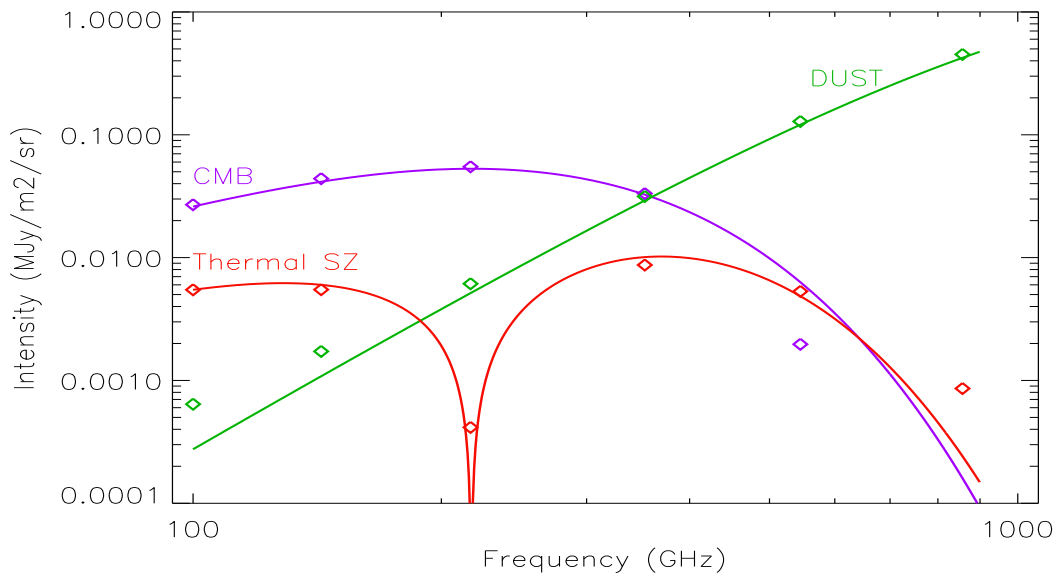


recovered SZ



Estimations des spectres électromagnétiques

☐ Spectres électromagnétiques



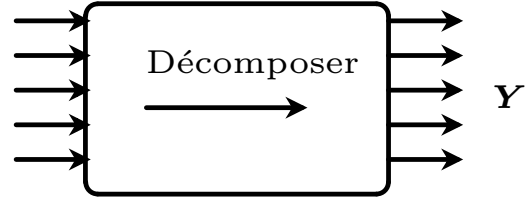
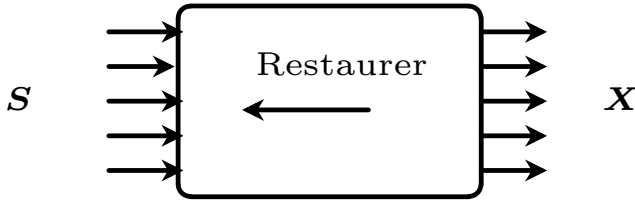
☐ Implémentation dans le domaine des **harmoniques sphériques** pour les données tout ciel.

Conclusions / Perspectives

Conclusions

- ❑ Importance de l'approche bayésienne :
 - Tenir compte du bruit dans le modèle.
 - Incorporer des informations *a priori* sur la matrice de mélange A → enlever des indéterminations.
 - Incorporer des informations sur la structure et les spectres des sources.
 - Cadre logique à l'incorporation de la ségmentation.
 - Considérer d'autres estimateurs que le maximum *a posteriori* grâce à l'échantillonnage bayésien.
- ❑ Caractérisation et élimination de la dégénérescence du maximum de vraisemblance.
- ❑ Sélection d'*a priori* et géométrie de l'information.
- ❑ Application en cosmologie observationnelle.

Perspectives



Méthodologie bayésienne

$$x = As + b$$

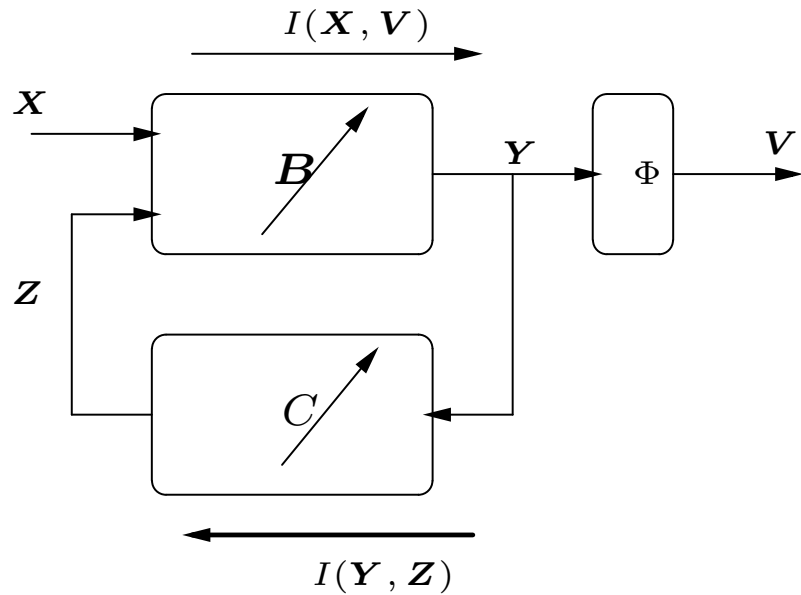
Calcul des probabilités

$$p(S, Z, A | X) \propto p(X | S, A) p(S | Z) p(Z) p(A) p(\theta)$$

↓
Traitement conjoint

Théorie de l'information

$$v = \Phi(Bx)$$



↓
Traitement séquentiel

Espace des propositions ↔ Espace des questions

Vers la logique des questions...