Lien entre méthodes de séparation de sources, ACP, ACI et principes d'apprentissage

Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes

UMR 08506 du CNRS-Supélec-UPS

Supélec, Plateau de Moulon

91192 Gif-sur-Yvette, FRANCE.

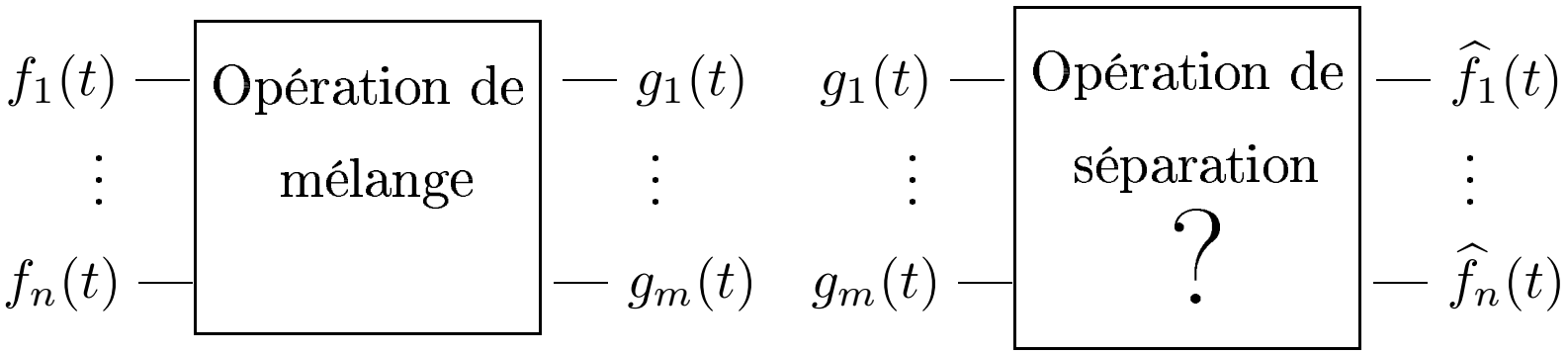
djafari@lss.supelec.fr

http ://www.lss.supelec.fr

http ://djafari.free.fr

Introduction

Problème de la séparation de source :



- **Mélange linéaire :**

$$\boldsymbol{g}(t) = \int \boldsymbol{A}(t, t')\boldsymbol{f}(t')\,\mathrm{d}t'$$

- **Mélange convolutive :**

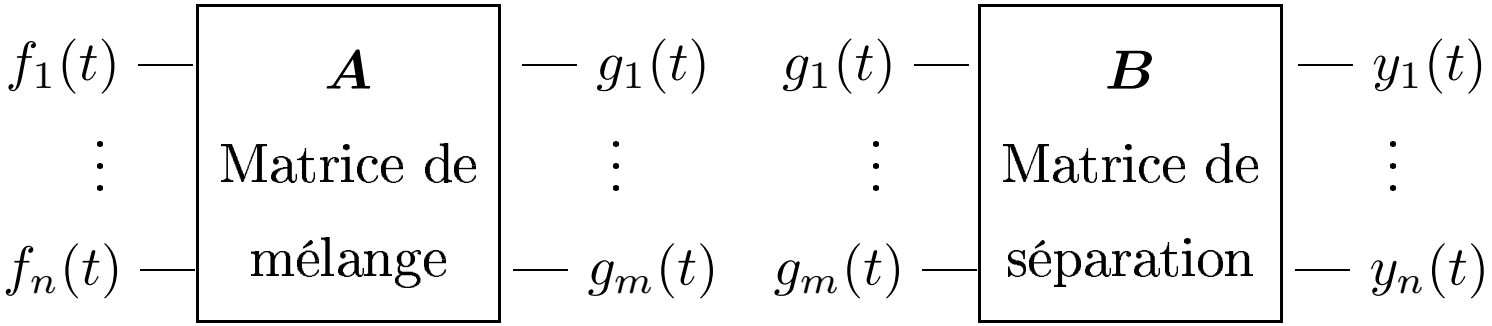
$$\boldsymbol{g}(t) = \int \boldsymbol{A}(t - t')\boldsymbol{f}(t')\,\mathrm{d}t'$$

- **Mélange instantanée :**

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t)$$

Mélange convolutive $\longrightarrow$ **Déconvolution multi canaux**

Mélange instantanée $\longrightarrow$ **Séparation de sources**



- **Indétermination principale :** $\longrightarrow \boldsymbol{B} = \boldsymbol{\Sigma}\,\boldsymbol{\Lambda}\,\boldsymbol{A}^{-1}$
  où $\boldsymbol{\Sigma}$ est une matrice de permutation et $\boldsymbol{\Lambda}$ une matrice diagonale.
- Hypothèses fondamentales : $f_1(t), \ldots, f_n(t)$ sont non corrélés (ACP) ou independants (ACI).
- Méthodes classiques : Infomax, Contrast, Ordres supérieures, Maximum de vraisemblance, Approche bayésienne

Analyse en composants principales (ACP)

- **Principale hypothèse :**

  $f_1(t), \ldots, f_n(t)$ sont non corrélés entre eux et blanches.

$$\boldsymbol{R}_{ff} = \mathrm{E}\,\boldsymbol{f}\boldsymbol{f}^t = \boldsymbol{\Lambda}$$

- **Algorithme :** $\boldsymbol{R}_{gg} = \mathrm{E}\,\boldsymbol{g}\boldsymbol{g}^t = \mathrm{E}\,\boldsymbol{A}\boldsymbol{f}\boldsymbol{f}^t\boldsymbol{A}^t = \boldsymbol{A}\boldsymbol{R}_{ff}\boldsymbol{A}^t = \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^t$
  - Estimation de la matrice de covariance

$$[\boldsymbol{R}_{gg}]_{kl} = \frac{1}{T}\sum_t g_k(t)\, g_l(t)$$

  - Décomposition en valeurs singulières

$$\boldsymbol{R}_{gg} = \boldsymbol{A}\Lambda\boldsymbol{A}^t \longrightarrow \widehat{\boldsymbol{f}}(t) = (\Lambda^+)^{1/2}\boldsymbol{A}^t\boldsymbol{g}(t)$$

- $\boldsymbol{A}$ est déterminée à une rotation et à un facteur d'échelle près :

$$\boldsymbol{A} \longrightarrow \boldsymbol{A}\boldsymbol{\Theta} \quad \longrightarrow \quad \boldsymbol{A}\boldsymbol{\Theta}\Lambda\boldsymbol{\Theta}^t\boldsymbol{A}^t = \boldsymbol{A}\Lambda\boldsymbol{A}^t \quad \text{avec} \quad \boldsymbol{\Theta} \quad \text{une matrice orthogonale}$$

Analyse en composants indépendants (ACI)

– Imposer une structure pour l'estimateur :

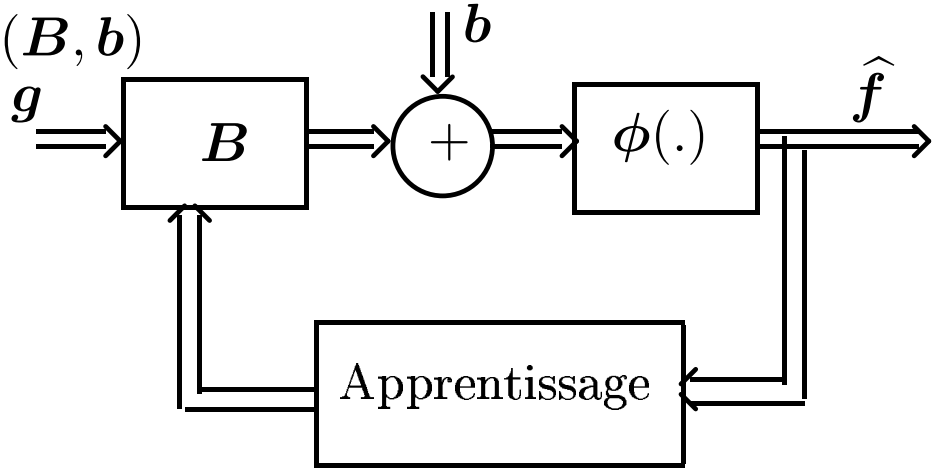
$$\widehat{f}_i = \phi_i([\boldsymbol{B}\boldsymbol{g} + \boldsymbol{b}]_i) \longrightarrow \widehat{\boldsymbol{f}} = \boldsymbol{\phi}(\boldsymbol{B}\boldsymbol{g} + \boldsymbol{b})$$

– Choisir un critère de mesure de l'indépendance des composants de $\widehat{\boldsymbol{f}}$:

$$S = -\sum_i p_i(\widehat{f}_i) \ln p_i(\widehat{f}_i) = -\sum_i p_i(\phi_i([\boldsymbol{B}\boldsymbol{g}]_i + b_i)) \ln p_i(\phi_i([\boldsymbol{B}\boldsymbol{g}]_i + b_i))$$

– Optimiser S : $(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{b}}) = \arg\max (\boldsymbol{B}, \boldsymbol{b}) S(\boldsymbol{B}, \boldsymbol{b})$

– Structure d'optimisation
dans un réseau de neurons



Méthodes basées sur des fonctions de contrasts

Définir une fonction de contraste $c(\boldsymbol{y}) = c(\boldsymbol{Bg})$ qui aura pour minimiseur la matrice de séparation $\boldsymbol{B}$.
Exemple :

$$c(\boldsymbol{B}) = KL\left(p(\boldsymbol{y}) : \prod_i p_i(y_i)\right) = \int p(\boldsymbol{y}) \ln \frac{p(\boldsymbol{y})}{\prod_i p_i(y_i)} \, \mathrm{d}\boldsymbol{y}$$

Méthodes basées sur les statistiques d'ordre supérieurs à deux

Approximation de $p_i(y_i)$ par ses cumulants

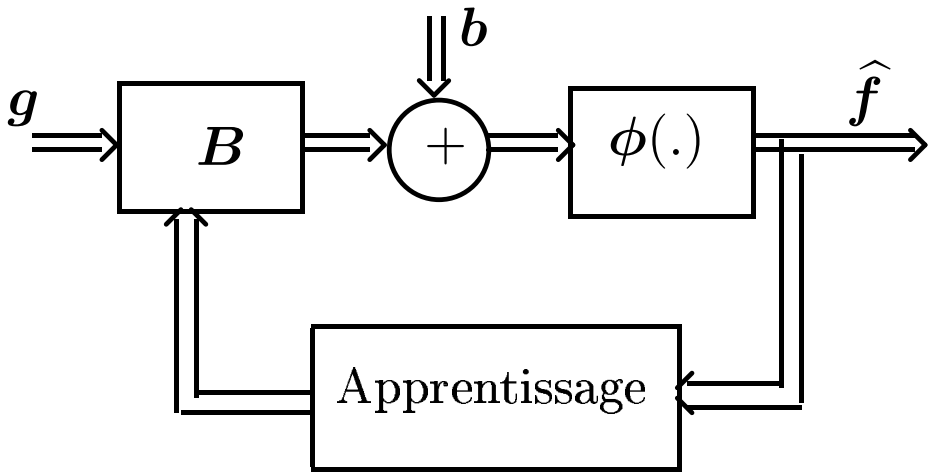
Dévélopement en série de Taylor de la fonction caractéristique :

$$\frac{\partial c(\boldsymbol{B})}{\partial \boldsymbol{B}} \propto \mathrm{E}\, \frac{p_i'(y_i)}{p_i(y_i)} \propto \text{cumulants de } (y_i)$$

Principales limitations des méthodes classiques

- La plus part de ces méthodes ne prennent pas en compte explicitement le bruit (les erreurs de la modélisation et le bruit de mesure) ;
- La plus part de ces méthodes supposent que la matrice de mélange est carrée et inversible.
- La plus part de ces méthodes supposent que les sources sont i.i.d. $\longrightarrow$ Balnchiement avant ACP ou ACI
- La plus part de ces méthodes imposent *a priori* une structure :

$$\widehat{\boldsymbol{f}} = \boldsymbol{\phi}(\boldsymbol{B}\boldsymbol{g} + \boldsymbol{b})$$



Approche bayésienne

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) + \boldsymbol{\epsilon}(t), \quad t = 1, \cdots, T$$

- Information sur $\boldsymbol{\epsilon}$ $\longrightarrow$ $p\left(\boldsymbol{g}_{1..T}|\boldsymbol{A}, \boldsymbol{f}_{1..T}\right)$
- Information sur $\boldsymbol{f}(t)$ $\longrightarrow$ $p(\boldsymbol{f}_{1..T})$
- Information sur $\boldsymbol{A}$ $\longrightarrow$ $p(\boldsymbol{A})$
- Bayes :

$$p\left(\boldsymbol{A}, \boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T}\right) \propto p\left(\boldsymbol{g}_{1..T}|\boldsymbol{A}, \boldsymbol{f}_{1..T}\right)\ p(\boldsymbol{f}_{1..T})\ p(\boldsymbol{A})$$

- Définir un estimateur utilisant :

$$p\left(\boldsymbol{A}, \boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T}\right), \quad p\left(\boldsymbol{A}|\boldsymbol{g}_{1..T}\right) \quad \text{ou} \quad p(\boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T})$$

Exemples de $p(\boldsymbol{A})$:

$$p(\boldsymbol{A}) \propto \exp\left[-\lambda\|\boldsymbol{A}\|^2\right], \quad p(\boldsymbol{A}) \propto \exp\left[-\frac{1}{2\sigma_a^2}\|\boldsymbol{I} - \boldsymbol{A}\|^2\right]$$

Cas d'un modèle exact, A inversible et les sources iid

- ***A* inversible**

$$\boldsymbol{g}(t) = \boldsymbol{A}\boldsymbol{f}(t) = \boldsymbol{B}^{-1}\boldsymbol{f}(t), \quad t = 1, \ldots, T$$

- **Sources iid : $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$**

$$p(\boldsymbol{f}) = \prod_i p_i(f_i) \longrightarrow p(\boldsymbol{g}|\boldsymbol{B}) = |\det(\boldsymbol{B})| \prod_i p_i([\boldsymbol{B}\boldsymbol{x}]_i)$$

$$\ln p(\boldsymbol{g}_{1..T}|\boldsymbol{B}) = \ln |\det(\boldsymbol{B})|^T + \sum_t \sum_i p_i(y_i(t)), \quad \text{avec} \quad \boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{g}(t)$$

$$J(\boldsymbol{B}) = -\ln p(\boldsymbol{B}|\boldsymbol{g}_{1..T}) = -T \ln |\det(\boldsymbol{B})| - \sum_t \sum_i \ln p_i(y_i(t)) + \ln p(\boldsymbol{B}) + cte.$$

MAP ou Maximum de vraisemblance :

$$\frac{\partial J(\boldsymbol{B})}{\partial \boldsymbol{B}} = -\sum_t \boldsymbol{H}(\boldsymbol{y}(t)) \quad \text{avec} \quad H(\boldsymbol{y}) = \frac{\partial}{\partial \boldsymbol{B}} \left[\sum_i \ln p_i(y_i) + \ln |\det(\boldsymbol{B})| + \ln p(\boldsymbol{B}) \right]$$

Cas particular : $p(\boldsymbol{B})$ uniforme $\longrightarrow$ Maximum de vraisemblance

$$\boldsymbol{H}\left(\boldsymbol{y}\right) = \boldsymbol{\phi}(\boldsymbol{y})\,\boldsymbol{y}^{\dagger} - \boldsymbol{I},$$

avec

$$\phi_i(y_i) = -\frac{p_i'(y_i)}{p_i(y_i)}.$$

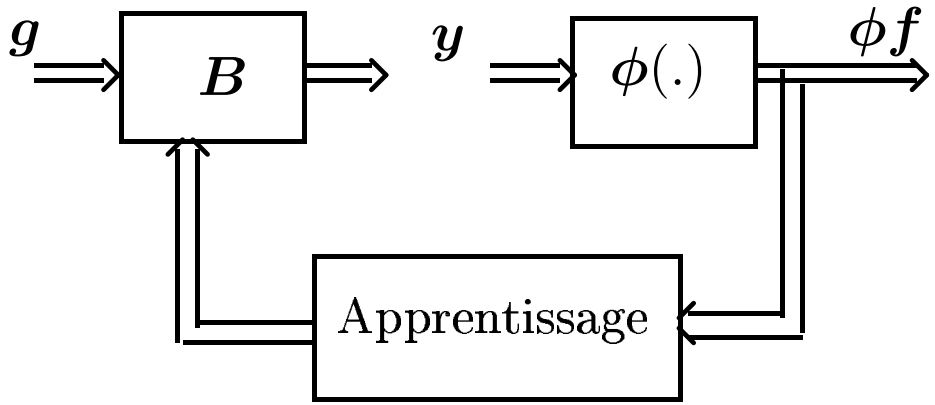
Gauss	$p(z) \propto \exp\left[-\alpha z^2\right]$	$\phi(z) = 2\alpha z$
Laplace	$p(z) \propto \exp\left[-\alpha\lvert z\rvert\right]$	$\phi(z) = \alpha\text{sign}(z)$
Cauchy	$p(z) \propto \frac{1}{1+(z/\alpha)^2}$	$\phi(z) = \frac{2z/\alpha^2}{1+(z/\alpha)^2}$
sous gaussien	$p(z) \propto \exp\left[-\frac{1}{2}z^2\right] \text{sech}^2(z)$	$\phi(z) = z + \tanh(z)$
mélange de gaussiennes	$p(z) \propto \exp\left[-\frac{1}{2}(z-\alpha)^2\right] + \exp\left[-\frac{1}{2}(z+\alpha)^2\right]$	$\phi(z) = \alpha z - \alpha\tanh(\alpha z)$

Lien avec l'apprentissage et RN

$$\boldsymbol{H}(\boldsymbol{y}) = \frac{\partial}{\partial \boldsymbol{B}} \left[\sum_i \ln p_i\left(y_i\right) - \ln \left|\det\left(\boldsymbol{B}\right)\right| \right].$$

Optimisation par un algorithme du descente :

$$\Delta \boldsymbol{B} \propto H(\boldsymbol{y}) = [\boldsymbol{I} - \boldsymbol{\phi}(\boldsymbol{y})\boldsymbol{y}^t]\boldsymbol{B}$$



Prise en compte des erreurs

$$\boldsymbol{g}(t) = \boldsymbol{A}\,\boldsymbol{f}(t) + \boldsymbol{\epsilon}(t), \quad t = 1, \ldots, T.$$

$$\ln p(\boldsymbol{\epsilon}(1), \ldots, \boldsymbol{\epsilon}(T)) = \sum_t \sum_i \ln p_i\left(\epsilon_i(t)\right).$$

- **Vraisemblence :**

$$\ln p\left(\boldsymbol{g}_{1..T} | \boldsymbol{A}, \boldsymbol{f}_{1..T}\right) = \sum_t \sum_i q_i\left(g_i(t) - [\boldsymbol{A}\boldsymbol{f}]_i(t)\right)$$

avec $q_i(.) = \ln p_i(.)$.

- **A posteriori :**

$$\begin{aligned} \ln p\left(\boldsymbol{A}, \boldsymbol{f}_{1..T} | \boldsymbol{g}_{1..T}\right) &= \ln p\left(\boldsymbol{g}_{1..T} | \boldsymbol{A}, \boldsymbol{f}_{1..T}\right) + \ln p(\boldsymbol{f}_{1..T}) + \ln p(\boldsymbol{A}) + cte \\ &= \sum_t \sum_i q_i\left(g_i(t) - [\boldsymbol{A}\boldsymbol{f}]_i(t)\right) + \ln p(\boldsymbol{f}_{1..T}) + \ln p(\boldsymbol{A}) + cte. \end{aligned}$$

Trois directions (dépendant de l'application) :

- Intégrer hors du problème $\boldsymbol{A}$ pour obtenir $p(\boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T})$ et estimer $\boldsymbol{f}_{1..T}$.
  Par exemple
  $$\widehat{\boldsymbol{f}}_{1..T} = \arg\max \boldsymbol{f}_{1..T} p(\boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T})$$
- Intégrer hors du problème $\boldsymbol{f}_{1..T}$ pour obtenir $p(\boldsymbol{A}|\boldsymbol{g}_{1..T})$ et estimer $\boldsymbol{A}$.
  Par exemple
  $$\widehat{\boldsymbol{A}} = \arg\max \boldsymbol{A} p(\boldsymbol{A}|\boldsymbol{g}_{1..T}),$$
  et ensuite estimer $\boldsymbol{f}$ par $\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{A}}\boldsymbol{g}$;
- Estimation jointe de $\boldsymbol{f}_{1..T}$ et de $\boldsymbol{A}$ en utilisant $p(\boldsymbol{A}, \boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T})$.
  Par exemple

$$\begin{cases} \widehat{\boldsymbol{f}}^{(k)}_{1..T} &= \arg\max \boldsymbol{f}_{1..T} p\left(\widehat{\boldsymbol{A}}^{(k-1)}, \boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T}\right) \\ \widehat{\boldsymbol{A}}^{(k)} &= \arg\max \boldsymbol{A} p\left(\boldsymbol{A}, \widehat{\boldsymbol{f}}^{(k-1)}_{1..T}|\boldsymbol{g}_{1..T}\right) \end{cases}$$

Sources indépendantes et blanches

$$\ln p(\boldsymbol{f}_{1..T}) = \sum_t \sum_j r_j(f_j(t))$$

$$p(\boldsymbol{A}) \propto \exp\left[-\frac{1}{2\sigma_a^2}\sum_k \sum_l a_{kl}^2\right],$$

$$\ln p\left(\boldsymbol{A}, \boldsymbol{f}_{1..T}|\boldsymbol{g}_{1..T}\right) = \sum_t \sum_i q_i\left(g_i(t) - y_i(t)\right) + \sum_t \sum_j r_j(f_j(t)) + \frac{1}{2\sigma_a^2}\sum_k \sum_l a_{kl}^2 + cte.$$

$$\begin{cases} \widehat{\boldsymbol{f}}^{(k)} &= \arg\max_{\boldsymbol{f}} \sum_i q_i\left(g_i - y_i\right) + \sum_j r_j(f_j) \\ \widehat{\boldsymbol{A}}^{(k)} &= \arg\max_{\boldsymbol{A}} \sum_i q_i\left(g_i - y_i\right) + \frac{1}{2\sigma_a^2}\sum_k \sum_l a_{kl}^2 \end{cases}$$

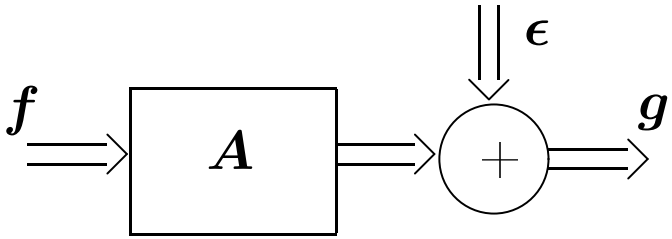
– **Lois gaussiennes** pour le bruit et les sources :

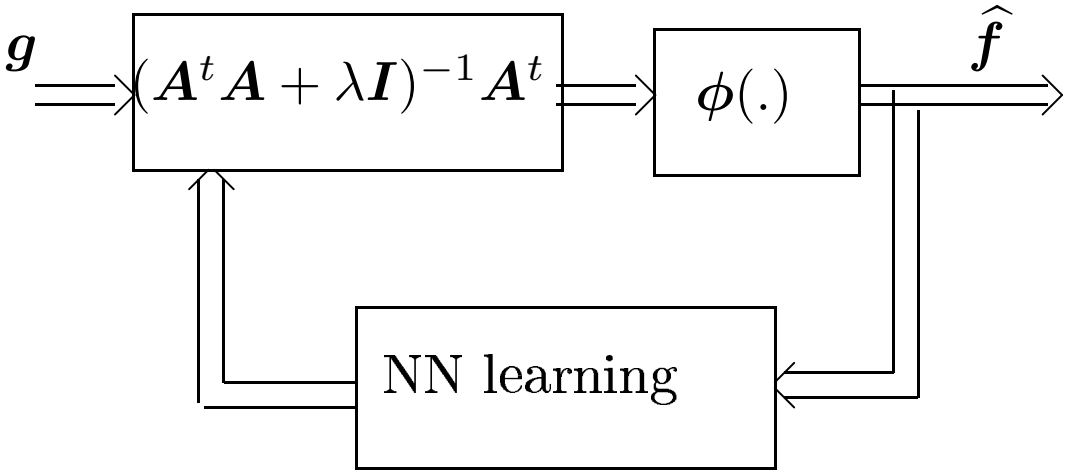
$$\begin{cases} \boldsymbol{f} &= (\boldsymbol{A}^t\boldsymbol{A}+\lambda\boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{g} \\ \boldsymbol{A} &= \boldsymbol{g}\boldsymbol{f}^t(\boldsymbol{f}\boldsymbol{f}^t+\mu\boldsymbol{I})^{-1} \end{cases}$$

avec $\lambda = \sigma_n^2/\sigma_s^2$ et $\mu = \sigma_n^2/\sigma_a^2$.

– **Lois non gaussienne pour $\boldsymbol{f}$:**

$$\begin{cases} \boldsymbol{y} &= (\boldsymbol{A}^t\boldsymbol{A}+\lambda\boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{g} \\ \boldsymbol{f} &= \boldsymbol{\phi}(\boldsymbol{y}) \\ \boldsymbol{A} &= \boldsymbol{g}\boldsymbol{f}^t(\boldsymbol{f}\boldsymbol{f}^t+\mu\boldsymbol{I})^{-1} \end{cases}$$





Sources dépendantes

$$\boldsymbol{f} = (\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{g} \quad \longrightarrow \quad \boldsymbol{f} = (\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{D}^t\boldsymbol{D})^{-1}\boldsymbol{A}^t\boldsymbol{g}$$

Sources indépendantes mais colorées temporellement

Principale difficultée : Modélisation de $p(f_j(1), \ldots, f_j(T))$

- **Modèles markoviens :**

$$\ln p(f_j(1), \ldots, f_j(T)) = \sum_t \ln p(f_j(t)|f_j(t-1))$$

- **Modèle de Gauss-Markov :**

$$\boldsymbol{f}(t) = (\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\left[\text{diag}\left[\lambda_1, \ldots, \lambda_n\right]\boldsymbol{f}(t-1) + \boldsymbol{A}^t\boldsymbol{g}(t)\right]$$

Marginalisation

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{\epsilon}$$

$$p(\boldsymbol{A}, \boldsymbol{f}|\boldsymbol{g}) \propto \exp\left[-\frac{1}{2\sigma_n^2} J(\boldsymbol{A}, \boldsymbol{f})\right] \quad \text{avec} \quad J(\boldsymbol{A}, \boldsymbol{f}) \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \lambda\phi(\boldsymbol{f}) + \mu\psi(\boldsymbol{A})$$

$$p(\boldsymbol{A}|\boldsymbol{g}) = \int p(\boldsymbol{A}, \boldsymbol{f}|\boldsymbol{g})\, \mathrm{d}\boldsymbol{f}$$

Approximation du second ordre :

$$-\ln p(\boldsymbol{A}|\boldsymbol{g}) \propto -\ln \left|\det\left(\widehat{\boldsymbol{P}}_s^{-1}\right)\right| - J(\boldsymbol{A}, \widehat{\boldsymbol{f}})$$

$$\widehat{\boldsymbol{f}} = (\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{g} \quad \text{and} \quad \widehat{\boldsymbol{P}}_s = (\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{I})^{-1};$$

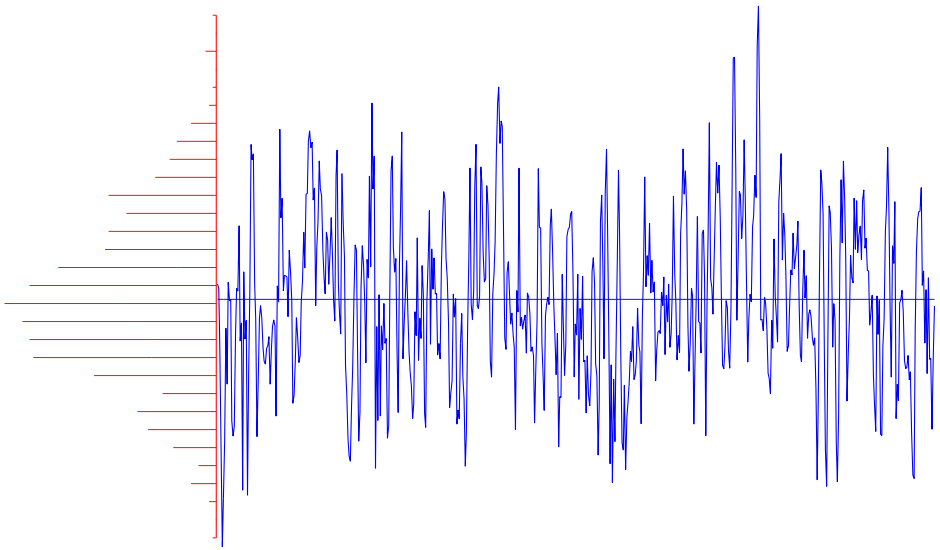
$\longrightarrow$ On obtient le même schéma algorithmique de RN avec :

$$\Delta\boldsymbol{A} \propto \boldsymbol{A}^t(\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{I})^{-1} + \boldsymbol{g}\boldsymbol{f} + \mu\boldsymbol{\psi}'(\boldsymbol{A})$$

Algorithme Espérance-Maximisation (EM) :

Données complètes $(\boldsymbol{g}, \boldsymbol{f})$ et incomplètes $\boldsymbol{g}$.

Choix d'une loi de probabilité a priori

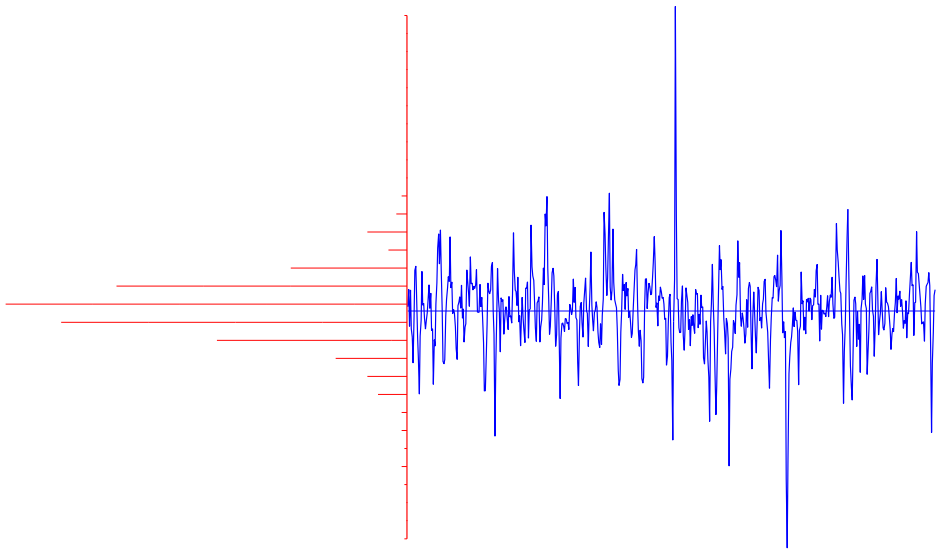


Gaussienne (Gauss-Markov)

$$p(f_j|f_{j-1}) = \mathcal{N}(f_{j-1}, \sigma_f^2)$$

$$p(\boldsymbol{f}) \propto \exp\left[-\alpha \sum_j |f_j - f_{j-1}|^2\right]$$

$$p(\boldsymbol{f}) \propto \exp\left[-\alpha \sum_{\boldsymbol{r}\in\mathcal{R}} \left|f(\boldsymbol{r}) - \beta \sum_{\boldsymbol{s}\in\mathcal{V}(\boldsymbol{r})} f(\boldsymbol{s})\right|^2\right]$$



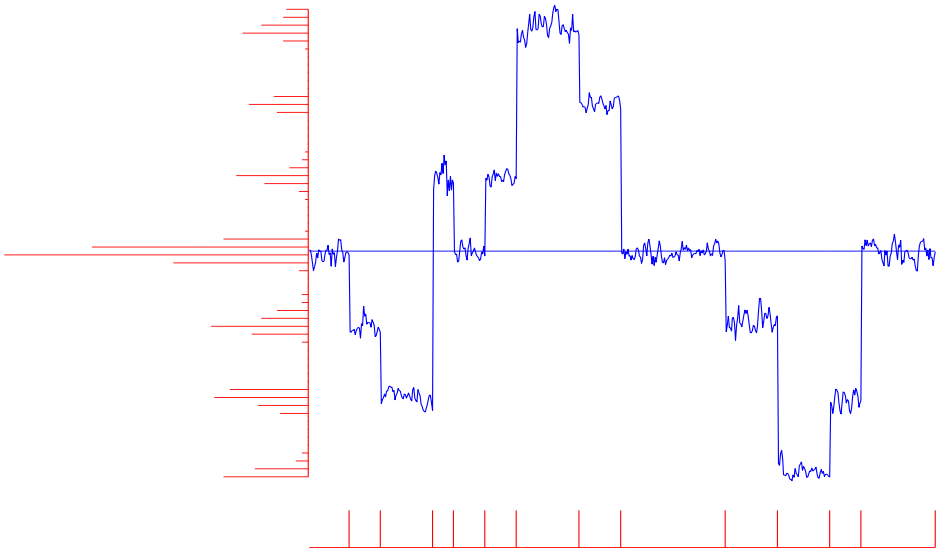
Gaussienne généralisée

$$p(f_j|f_{j-1}) \propto \exp\left[-\alpha|f_j - f_{j-1}|^p\right]$$

$$p(\boldsymbol{f}) \propto \exp\left[-\alpha \sum_j |f_j - f_{j-1}|^p\right]$$

$$p(\boldsymbol{f}) \propto \exp\left[-\alpha \sum_{\boldsymbol{r}\in\mathcal{R}} \left|f(\boldsymbol{r}) - \beta \sum_{\boldsymbol{s}\in\mathcal{V}(\boldsymbol{r})} f(\boldsymbol{s})\right|^p\right]$$

Modèles a priori hiérarchiques avec champs cachés (HMRF)

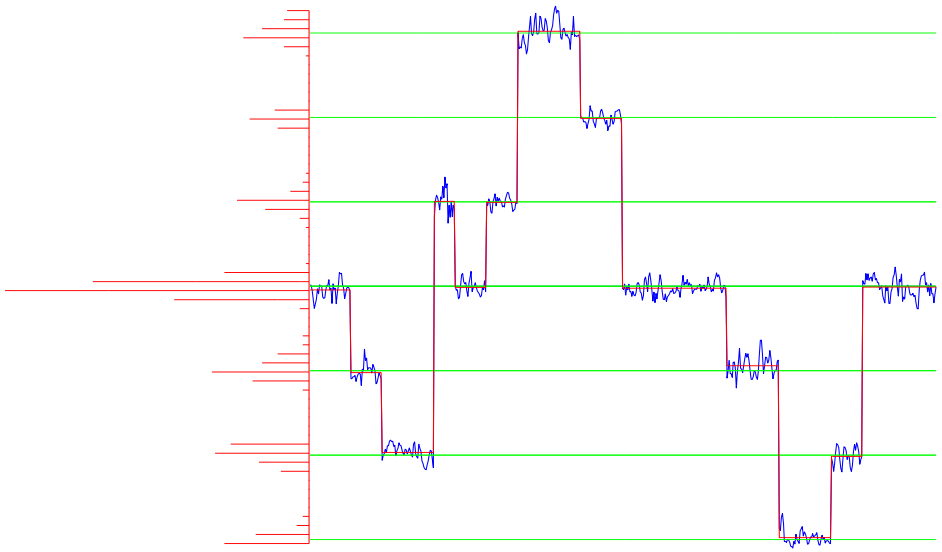


Gauss-Markov + processus de ligne

$$p(f_j|q_j, f_{j-1}) = \mathcal{N}\left((1-q_j)f_{j-1}, \sigma_f^2\right)$$

$$p(\boldsymbol{f}|\boldsymbol{q}) \propto \exp\left[-\alpha \sum_j (1-q_j)|f_j - f_{j-1}|^2\right]$$

$$P(q_j = 1) = \alpha, P(q_j = 0) = 1-\alpha$$



Mélange de gaussiennes ou GM+régions

$$p(f_j|z_j = k) = \mathcal{N}\left(m_k, \sigma_k^2\right)$$

$$p(\boldsymbol{f}|\boldsymbol{z}) \propto \exp\left[-\alpha \sum_k \sum_{j \in \mathcal{R}_k} (f_j - m_k/\sigma_k)^2\right]$$

$$p(\boldsymbol{z}) \propto \exp\left[\alpha \sum_j \sum_{i \in \mathcal{V}(j)} \delta(z_j - z_i)\right]$$

Modélisation mélange de gaussiennes et variables cachées

$$p(f_j|z_j=k)=\mathcal{N}(m_{jk},\sigma^2_{jk}) \longrightarrow p(f_j)=\sum_{k=1}^{K_j}\alpha_{jk}\mathcal{N}(m_{jk},\sigma^2_{jk})$$

- **Variable cachée :** $z_j \in \mathcal{Z}_j=(1,\ldots,K_j)$ avec $\alpha_{jk}=p(z_j=k)$

$$p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta})=\prod_j\prod_{z_j\in\mathcal{Z}_j}\mathcal{N}(m_{jk},\sigma^2_{jk}), \qquad \boldsymbol{\theta}=\{(m_{jk},\sigma^2_{jk})\}$$

  - Loi *a priori* $\quad p(\boldsymbol{f}|\boldsymbol{\theta})=\sum_{\boldsymbol{z}\in\mathcal{Z}} p(\boldsymbol{z})\,p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta})$
  - Loi *a posteriori* $\quad p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta})=\sum_{\boldsymbol{z}\in\mathcal{Z}} p(\boldsymbol{z}|\boldsymbol{g},\boldsymbol{\theta})\,p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{z},\boldsymbol{\theta})$

- **Algorithme globale de l'estimation :**
  - Estimer $\boldsymbol{z}$ utilisant $p(\boldsymbol{z}|\boldsymbol{g},\boldsymbol{\theta})$;
  - Estimer $\boldsymbol{f}$ utilisant $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{z},\boldsymbol{\theta})$
  - Estimer les hyperparamèters $\boldsymbol{\theta}$ utilisant $p(\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{z},\boldsymbol{f})$

Conclusions

- Il y a un lien entre les problèmes de séparation de sources et la notion de l'apprentissage.
- Les méthodes classiques supposent $\boldsymbol{A}$ inversible et ne prennent pas en compte explicitement les incertitudes.
- Approche bayésienne permet de combiner l'information dans les données et l'information *a priori*, et prends en compte explicitement les incertitudes.
- Dans les méthodes classique la structure du schéma de l'apprentissage est fixée d'avance. Dans l'approche bayésienne, cette structure dépend de la modélisation du bruit et des signaux d'entrée.
- En général, le calcul bayésien est coûteux :
  nécessité de faire des approximations,
  par exemple l'approximation en champs moyens (Mean field approximation)