Analyse Factorielle parcimonieuse avec priors Bernoulli-Gaussiens sur les facteurs

Décembre 2018

I. Modèle 1 : Parcimonie par priors exponentiels

L'objectif est d'améliorer le modèle suivant :

$$\boldsymbol{y}_i = \boldsymbol{L}\lceil \boldsymbol{q} \rfloor \boldsymbol{c}_i + \boldsymbol{n}_i \tag{1}$$

Priors	**Hyperpriors**
$\boldsymbol{L} \sim \mathcal{N}_c(0, \frac{\boldsymbol{I}_{MK}}{K})$	
$\boldsymbol{c} \sim \mathcal{N}_c(0, \lceil \boldsymbol{\gamma^2} \rfloor)$	$\boldsymbol{\gamma^2} \sim \mathcal{IG}(a_\gamma, b_\gamma)$
$\boxed{\boldsymbol{q} \sim \mathcal{E}xp(\boldsymbol{a}_q)}$	
$\boldsymbol{n}\lvert\beta^2 \sim \mathcal{N}_c(0, \beta^2 \lceil \boldsymbol{\sigma^2} \rfloor)$	$\boldsymbol{\sigma^2} \sim \mathcal{IG}(a_\sigma, b_\sigma)$ $\beta^2 \sim \mathcal{IG}(a_\beta, b_\beta)$

Ce modèle pose les problèmes résumés dans le tableau ci-dessous :

Faiblesses rencontrées dans le modèle 1	**Solutions proposées par le modèle 2**
$\boldsymbol{q}$ n'est pas normalisé, on ne peut donc pas observer la stabilité de la chaîne	$\boldsymbol{q}$ vaut 0 ou 1 ; toute la variance est portée par $\boldsymbol{c}$
$\boldsymbol{q}$ et $\boldsymbol{c}$ risquent d'être fortement corrélés, ce qui peut nuire à la mélangeance des chaînes	Marginalisation partielle de l'échantillonneur
Sensibilité des résultats au choix des hyperparamètres $\boldsymbol{a}_q$	Échantillonnage de l'hyperparamètre dans une loi Bêta.

II. Bibliographie

Le chapitre 4 de la thèse de Charly Faure y est dédié, avec le modèle suivant :

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{n} \tag{2}$$

où $\boldsymbol{f} \sim \prod_i \mathcal{B}ern\mathcal{G}auss(f_i\ ;\ l, \gamma^2)$. L'échantillon f_i est donc tiré soit dans une Gaussienne, soit dans un Dirac en 0 (cas particulier d'une Gaussienne de variance nulle). Ce qui peut s'écrire :

$$\mathcal{B}ern\mathcal{G}auss(l, \gamma^2) = (1-l)\mathcal{N}_c(0, 0) + l\mathcal{N}_c(0, \gamma^2) \tag{3}$$

Ge et al. (2011) ont une approche similaire, en notant $\boldsymbol{f} \sim \mathcal{N}_c(0, \lceil \boldsymbol{q} \rfloor \gamma^2)$, où les éléments de $\boldsymbol{q}$ suivent une loi de Bernoulli.

3.1. Nomenclature

M	Nombre de micros
K	Nombre de facteurs
I_s	Nombre de moyennes pour le calculs des CSM
i	indice associé à une moyenne
L_q	nombre d'éléments non-nuls du vecteur $\boldsymbol{q}$

3.2. Modèle 2 : Parcimonie par priors bernoulliens

On choisit un modèle hiérarchique avec la matrice de mixage $\boldsymbol{L}$, les facteurs $\boldsymbol{c}$ et les binaires $\boldsymbol{q}$ au même étage :

$$\boldsymbol{y}_i = \boldsymbol{L}\lceil \boldsymbol{q} \rfloor \boldsymbol{c}_i + \boldsymbol{n}_i \tag{4}$$

Priors	**Hyperpriors**
$\boldsymbol{L} \sim \mathcal{N}_c(0, \frac{\boldsymbol{I}_{MK}}{K})$	
$\boldsymbol{c} \sim \mathcal{N}_c(0, \lceil \boldsymbol{\gamma^2} \rfloor)$	$\boldsymbol{\gamma^2} \sim \mathcal{IG}(a_\gamma, b_\gamma)$
$\boxed{\boldsymbol{q} \sim \mathcal{B}ern(l)}$	$\boxed{l \sim \mathcal{B}eta(a_l, b_l)}$
$\boldsymbol{n}\|\beta^2 \sim \mathcal{N}_c(0, \beta^2\lceil \boldsymbol{\sigma^2} \rfloor)$	$\boldsymbol{\sigma^2} \sim \mathcal{IG}(a_\sigma, b_\sigma)$ $\beta^2 \sim \mathcal{IG}(a_\beta, b_\beta)$

On note par la suite $\boldsymbol{\Omega}_n = \beta^2\lceil \boldsymbol{\sigma^2} \rfloor$.

3.3. Posteriors

3.3.1 Échantillonnage du paramètre de parcimonie l

$$\begin{aligned}
[l|\boldsymbol{q}] &\propto \mathcal{B}eta(l; a_l, b_l) \prod_k^K \mathcal{B}ern(q_k; l) \\
&\propto \frac{\Gamma(a_l + b_l)}{\Gamma(a_l)\Gamma(b_l)} l^{a_l - 1}(1-l)^{b_l - 1} \prod_k^K l^{q_k}(1-l)^{1-q_k} \\
&\propto \frac{\Gamma(a_l + b_l)}{\Gamma(a_l)\Gamma(b_l)} \; l^{a_l - 1} \; (1-l)^{b_l - 1} \; l^{\sum_k q_k} \; (1-l)^{\sum_k 1 - q_k} \\
&\boxed{\propto \mathcal{B}eta(a_l + L_q, b_l + K - L_q)}
\end{aligned}$$

3.3.2 Échantillonnage en bloc de c

$$
\begin{aligned}
[\boldsymbol{c}_i|\infty_{-\boldsymbol{c}_i}] &\propto [\boldsymbol{y}_i|\infty_{-\boldsymbol{y}_i}]\,\mathcal{N}_c(\boldsymbol{c}_i;0,\boldsymbol{q}\boldsymbol{\gamma}^2)\\
&\propto \frac{\mathrm{e}^{-(\boldsymbol{y}_i-\boldsymbol{L}\boldsymbol{q}\boldsymbol{c}_i)^H\boldsymbol{\Omega}_n^{-1}(\boldsymbol{y}_i-\boldsymbol{L}\boldsymbol{q}\boldsymbol{c}_i)}}{|\boldsymbol{\Omega}_n|}\mathrm{e}^{\boldsymbol{c}_i^H\boldsymbol{\gamma}^{-2}\boldsymbol{c}_i}\\
&\propto \mathrm{e}^{-(\boldsymbol{c}_i-\boldsymbol{\mu}_i)^H\boldsymbol{\Omega}_{\boldsymbol{c}_i}^{-1}(\boldsymbol{c}_i-\boldsymbol{\mu}_i)}
\end{aligned}
$$

Par identification, on a :

- $\boldsymbol{c}_i^H\boldsymbol{\Omega}_{\boldsymbol{c}_i}^{-1}\boldsymbol{c}_i = \boldsymbol{c}_i^H\left(\boldsymbol{q}\boldsymbol{L}^H\boldsymbol{\Omega}_n^{-1}\boldsymbol{L}\boldsymbol{q}+\boldsymbol{\gamma}^{-2}\right)\boldsymbol{c}_i$

- $\boldsymbol{\Omega}_{\boldsymbol{c}_i}^{-1}\boldsymbol{\mu}_i = \boldsymbol{q}^H\boldsymbol{L}^H\boldsymbol{\Omega}_n^{-1}\boldsymbol{y}_i$

Finalement :

$$
\boxed{
\begin{gathered}
[\boldsymbol{c}_i|\infty_{-\boldsymbol{c}_i}] \propto \mathcal{N}_c(\boldsymbol{\mu}_i,\boldsymbol{\Omega}_{\boldsymbol{c}_i}) \qquad (5)\\[2em]
\begin{aligned}
&\text{où :}\quad \boldsymbol{\Omega}_{\boldsymbol{c}_i} = \left(\boldsymbol{q}\boldsymbol{L}^H\boldsymbol{\Omega}_n^{-1}\boldsymbol{L}\boldsymbol{q}+\boldsymbol{\gamma}^{-2}\right)^{-1}\\
&\text{et}\quad \boldsymbol{\mu}_i = \boldsymbol{\Omega}_{\boldsymbol{c}_i}\boldsymbol{q}^H\boldsymbol{L}^H\boldsymbol{\Omega}_n^{-1}\boldsymbol{y}_i
\end{aligned}
\end{gathered}
}
$$

Dans le cas où le nombre de moyenne est suffisamment grand, on peut échantillonner directement la matrice interspectrale $\boldsymbol{S}_{\boldsymbol{cc}} = \mathbb{E}\{\boldsymbol{c}_i\boldsymbol{c}_i^H\}$. On a :

$$\boldsymbol{c}_i = \boldsymbol{\mu}_i + \boldsymbol{x}_i \quad \text{avec} \quad \boldsymbol{x}_i \sim \mathcal{N}_c(0,\boldsymbol{\Omega}_{\boldsymbol{c}_i}),$$

donc :

$$\mathbb{E}\{\boldsymbol{c}_i\boldsymbol{c}_i^H\} = \mathbb{E}\{\boldsymbol{\mu}_i\boldsymbol{\mu}_i^H\} + \mathbb{E}\{\boldsymbol{x}_i\boldsymbol{x}_i^H\} + \underbrace{2\mathbb{E}\{\boldsymbol{x}_i\boldsymbol{\mu}_i^H\}}_{\to 0 \text{ quand } I_s\to\infty},$$

avec :

- $\mathbb{E}\{\boldsymbol{x}_i\boldsymbol{x}_i^H\} = \boldsymbol{W}_c \sim \mathcal{W}(\boldsymbol{\Omega}_{\boldsymbol{c}_i}, I_s)$
- $\mathbb{E}\{\boldsymbol{\mu}_i\boldsymbol{\mu}_i^H\} = \boldsymbol{\Omega}_{\boldsymbol{c}_i}\boldsymbol{q}^H\boldsymbol{L}^H\boldsymbol{\Omega}_n^{-1}\boldsymbol{S}_{yy}\boldsymbol{L}\boldsymbol{q}\boldsymbol{\Omega}_{\boldsymbol{c}_i}^H$

3.3.3 Échantillonnage de $\boldsymbol{L}$

La matrice $\boldsymbol{L}$ est vectorisée de manière à écrire une matrice de covariance pour la postérieure qui soit à 2 dimensions. On note $\boldsymbol{\lambda} = \mathrm{vec}(\boldsymbol{L})$

$$[\boldsymbol{\lambda}|\infty_{\boldsymbol{\lambda}}] \propto \prod_i^{I_s}[\boldsymbol{y}_i|\infty_{-\boldsymbol{y}_i}][\boldsymbol{\lambda}]$$

Or :

$$\mathrm{vec}(\boldsymbol{y}_i) = \boldsymbol{y}_i = \mathrm{vec}(\boldsymbol{L}\boldsymbol{q}\boldsymbol{c}_i) + \boldsymbol{n}_i = \left(\boldsymbol{c}_i^T\boldsymbol{q}\otimes\boldsymbol{I}_M\right)\boldsymbol{\lambda} + \boldsymbol{n}_i, \qquad (6)$$

donc :

$$
\begin{aligned}
[\boldsymbol{\lambda}|\infty_{\boldsymbol{\lambda}}] &\propto \mathrm{e}^{-\sum_i \left(\boldsymbol{y}_i - \left(\boldsymbol{c}_i^T \boldsymbol{q} \otimes \boldsymbol{I}_K\right)\boldsymbol{\lambda}\right)^H \boldsymbol{\Omega}_n^{-1} \left(\boldsymbol{y}_i - \left(\boldsymbol{c}_i^T \boldsymbol{q} \otimes \boldsymbol{I}_M\right)\boldsymbol{\lambda}\right)} \mathrm{e}^{-\boldsymbol{\lambda}^H K \boldsymbol{I}_{MK} \boldsymbol{\lambda}} \\
&\propto \mathrm{e}^{-(\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda)^H \boldsymbol{\Omega}_{\boldsymbol{\lambda}}^{-1} (\boldsymbol{\lambda}-\boldsymbol{\mu}_\lambda)}
\end{aligned}
$$

Comme précédemment, par identification, on a finalement :

$$
\boxed{
\begin{aligned}
&[\boldsymbol{\lambda}|\infty_{\boldsymbol{\lambda}}] \propto \mathcal{N}_c(\boldsymbol{\mu}_\lambda, \boldsymbol{\Omega}_{\boldsymbol{\lambda}}) \qquad (7) \\
&\text{où :} \quad \boldsymbol{\Omega}_{\boldsymbol{\lambda}} = \left(\sum_i (\boldsymbol{c}_i^T \boldsymbol{q} \otimes \boldsymbol{I}_M)^H \boldsymbol{\Omega}_n^{-1} (\boldsymbol{c}_i^T \boldsymbol{q} \otimes \boldsymbol{I}_M) + K \boldsymbol{I}_{MK}\right)^{-1} \\
&\qquad\qquad = \left((\boldsymbol{q} \boldsymbol{S}_{cc}^* \boldsymbol{q}) \otimes (\boldsymbol{\Omega}_n^{-1}) + K \boldsymbol{I}_{MK}\right)^{-1} \\
&\quad \text{et} \quad \boldsymbol{\mu}_\lambda = \boldsymbol{\Omega}_{\boldsymbol{\lambda}} \sum_i (\boldsymbol{c}_i^T \boldsymbol{q} \otimes \boldsymbol{I}_M)^H \boldsymbol{\Omega}_n^{-1} \boldsymbol{y}_i \\
&\qquad\qquad = \boldsymbol{\Omega}_{\boldsymbol{\lambda}} \operatorname{vec}(\boldsymbol{\Omega}_n^{-1} \boldsymbol{S}_{yy} \boldsymbol{\Omega}_n^{-1} \boldsymbol{L} \boldsymbol{q} \boldsymbol{\Omega}_{\boldsymbol{c_i}} \boldsymbol{q})
\end{aligned}
}
$$

Notes : - Ce dernier résultats s'obtient en remplaçant $\boldsymbol{c}_i$ par $\boldsymbol{\mu}_i$.
- Les propriétés du produit de Kronecker sont les suivantes :

$$
\begin{aligned}
\operatorname{vec}(ABC) &= (C^T \otimes A) \operatorname{vec}(B) \\
(A \otimes B)^T &= A^T \otimes B^T \\
(A \otimes B)^* &= A^* \otimes B^* \\
(A \otimes B)(C \otimes D) &= (AC) \otimes (BD)
\end{aligned}
$$

si les dimensions des matrices A, B, C et D le permettent.

3.3.4 Échantillonnage de $\boldsymbol{\gamma}^2$

Cas où $\boldsymbol{\gamma}^2 = \gamma^2 \boldsymbol{I}_K$

$$
\begin{aligned}
\left[\gamma^2|\boldsymbol{S}_{cc}\right] &\propto \prod_i [\boldsymbol{c}_i|\gamma^2][\gamma^2] \\
&\propto \frac{\mathrm{e}^{-\gamma^{-2}\sum_i \boldsymbol{c}_i^H \boldsymbol{c}_i}}{\gamma^{2 I_s K}} \frac{\mathrm{e}^{-\gamma^{-2} b_\gamma}}{\gamma^{2(a_\gamma - 1)}}
\end{aligned}
$$

$$
\boxed{\left[\gamma^2|\boldsymbol{S}_{cc}\right] \propto \mathcal{IG}\left(a_\gamma + K I_s, b_\gamma + \operatorname{Trace}(\boldsymbol{S}_{cc})\right)} \qquad (8)
$$

Cas où $\boldsymbol{\gamma}^2 = \lceil \boldsymbol{\gamma}^2 \rfloor = \operatorname{diag}(\gamma_1^2, \ldots, \gamma_k^2, \ldots, \gamma_K^2)$

$$
\left[\gamma_k^2|\boldsymbol{S}_{cc}\right] \propto \prod_i [\boldsymbol{c}_{ik}|\gamma_k^2][\gamma_k^2]
$$

$$
\boxed{\left[\gamma_k^2|\boldsymbol{S}_{cc}\right] \propto \mathcal{IG}\left(a_\gamma + I_s, b_\gamma + \boldsymbol{S}_{cc_{kk}}\right)} \qquad (9)
$$

3.3.5 Échantillonnage de $\boldsymbol{\sigma}^2$

On considère un bruit hétéroskédastique : $\boldsymbol{\sigma}^2 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2, \ldots, \sigma_M^2)$.

$$\begin{aligned}
\left[\boldsymbol{\sigma}^2 | \infty_{-\boldsymbol{\sigma}^2}\right] &\propto \prod_i [\boldsymbol{y_i} | \infty_{-\boldsymbol{y}_i}][\boldsymbol{\sigma}] \\
&\propto \frac{\mathrm{e}^{-\sum_i (\boldsymbol{y}_i - \boldsymbol{Lqc}_i)^H \beta^{-2} \boldsymbol{\sigma}^{-2} (\boldsymbol{y}_i - \boldsymbol{Lqc}_i)}}{\beta^2 |\boldsymbol{\sigma}|^{I_s}} \frac{\mathrm{e}^{-\boldsymbol{\sigma}^{-2} b_\sigma}}{\boldsymbol{\sigma}^{2(a_\sigma - 1)}} \\
&\propto \frac{\mathrm{e}^{-\mathrm{Trace}\left(\beta^{-2}\boldsymbol{\sigma}^{-2} \overbrace{\sum_i \left\{\boldsymbol{y}_i \boldsymbol{y}_i^H + \boldsymbol{Lqc}_i \boldsymbol{c}_i^H \boldsymbol{qL}^H - \boldsymbol{y}_i \boldsymbol{c}_i^H \boldsymbol{qL}^H - \boldsymbol{Lqc}_i \boldsymbol{y}_i^H\right\}}^{\boldsymbol{T}}\right)}}{|\boldsymbol{\sigma}^2|^{I_s}} \frac{\mathrm{e}^{-\boldsymbol{\sigma}^{-2} b_\sigma}}{\boldsymbol{\sigma}^{2(a_\sigma - 1)}}
\end{aligned}$$

On réinjecte $\boldsymbol{c}_i = \boldsymbol{\mu}_i + \boldsymbol{x}_i$ et on suppose que $\mathbb{E}(\boldsymbol{x}_i \boldsymbol{y}_i^H) \approx 0$. En notant :

$$\boldsymbol{P} = \boldsymbol{Lq\Omega}_{c_i} \boldsymbol{qL}^H \boldsymbol{\Omega}_n^{-1} = \boldsymbol{P}^H,$$

on a : $\boldsymbol{Lq\mu}_i = \boldsymbol{Py}_i$. Finalement :

$$\begin{aligned}
\boldsymbol{T} &= \boldsymbol{S}_{yy} + \boldsymbol{LqW}_c \boldsymbol{qL}^H + \boldsymbol{PS}_{yy}\boldsymbol{P} - \boldsymbol{S}_{yy}\boldsymbol{P} - \boldsymbol{PS}_{yy} \\
&= (\boldsymbol{I}_M - \boldsymbol{P})\boldsymbol{S}_{yy}(\boldsymbol{I}_M - \boldsymbol{P}) + \boldsymbol{LqW}_c \boldsymbol{qL}^H
\end{aligned}$$

$$\boxed{\left[\sigma_m^2 | \infty_{-\sigma_m^2}\right] \propto \mathcal{IG}\left(a_\sigma + MI_s, b_\sigma + \boldsymbol{T}_{mm}\right)} \tag{10}$$

3.3.6 Échantillonnage de β^2

De manière similaire à l'échantillonnage de σ_m :

$$\begin{aligned}
\left[\beta^2 | \infty_{-\beta^2}\right] &\propto \prod_i [\boldsymbol{y_i} | \infty_{-\boldsymbol{y}_i}][\beta^2] \\
&\propto \frac{\mathrm{e}^{-\beta^{-2}\,\mathrm{Trace}(\boldsymbol{\sigma}^{-2}\boldsymbol{T})}}{|\beta^2 \boldsymbol{I}_M|^{I_s}} \frac{\mathrm{e}^{-\beta^{-2} b_\beta}}{\beta^{2(a_\beta - 1}}
\end{aligned}$$

$$\boxed{\left[\beta^2 | \infty_{-\beta^2}\right] \propto \mathcal{IG}\left(a_\beta + MI_s, b_\beta + \mathrm{Trace}(\boldsymbol{\sigma}^{-2}\boldsymbol{T})\right)} \tag{11}$$

3.3.7 Échantillonnage des binaires du vecteur $\boldsymbol{q}$

Sans marginalisation

$$\begin{aligned}
[q_k | \infty_{-q_k}] &\propto \prod_i \left[\boldsymbol{y}_i | q_k, \infty_{-\boldsymbol{y}_i}\right][q_k] \\
&\propto \frac{\mathrm{e}^{-\sum_i (\boldsymbol{y}_i - \boldsymbol{Lqc}_i)^H \boldsymbol{\Omega}_n^{-1} (\boldsymbol{y}_i - \boldsymbol{Lqc}_i)}}{|\boldsymbol{\Omega}_n|^{I_s}} l^{q_k} (1-l)^{1-q_k} \\
&\propto \mathrm{e}^{-\mathrm{Trace}\left(\boldsymbol{\Omega}_n^{-1}\boldsymbol{T}(q_k)\right) - q_k \ln\left(\frac{1}{l} - 1\right)} \\
&\propto \mathrm{e}^{-g(q_k)}
\end{aligned}$$

Le changement d'état d'un binaire se traduit par l'ajout de la quantité $\delta_k = (-1)^{q_k}$. On note $q_{k_{mod}} = q_k + \delta_k$. La probabilité que le binaire change d'état est :

$$\begin{aligned}
[q_{k_{mod}}|\infty_{-q_k}] &\propto \prod_i [\boldsymbol{y}_i|q_{k_{mod}}, \infty_{-\boldsymbol{y}_i,\boldsymbol{c}}]\, [q_{k_{mod}}] \\
&\propto \mathrm{e}^{-\operatorname{Trace}\left(\boldsymbol{\Omega}_n^{-1}\boldsymbol{T}(q_{k_{mod}})\right) - q_{k_{mod}} \ln\left(\frac{1}{l}-1\right)} \\
&\propto \mathrm{e}^{-g(q_{k_{mod}})}
\end{aligned}$$

Pour que la somme de ces deux probabilité soit égale à 1, il faut les normaliser. La probabilité qu'un binaire change d'état est alors :

$$\begin{aligned}
[q_{k_{mod}}|\infty_{-q_k}] &= \frac{\mathrm{e}^{-g(q_{k_{mod}})}}{\mathrm{e}^{-g(q_{k_{mod}})} + \mathrm{e}^{-g(q_k)}} \\
&\boxed{= \frac{1}{1 + \mathrm{e}^{-\left(g(q_k) - g(q_{k_{mod}})\right)}}}
\end{aligned}$$

Un échantillon $t \sim \mathcal{U}(0,1)$ est comparé à $[q_{k_{mod}}|\infty_{-q_k}]$. S'il y est inférieur, le binaire change d'état, sinon il conserve son état actuel.
En pratique, comme $\boldsymbol{T}$ peut prendre de grandes valeurs (not. si I_s est grand), le changement est accepté si

$$-\ln\left(\frac{1}{t} - 1\right) < g(q_k) - g(q_{k_{mod}})$$

Avec marginalisation sur les facteurs $\boldsymbol{c}$

$$[q_k|\infty_{-q_k,-\boldsymbol{c}}] \propto \prod_i [\boldsymbol{y}_i|q_k, \infty_{-\boldsymbol{y}_i,-\boldsymbol{c}}]\, [q_k]$$

Or,

$$[\boldsymbol{y}_i|q_k, \infty_{-\boldsymbol{y}_i,-\boldsymbol{c}}] = \mathcal{N}_c(0, \underbrace{\boldsymbol{L}\boldsymbol{q}\boldsymbol{\gamma}^2\boldsymbol{q}\boldsymbol{L}^H + \boldsymbol{\Omega}_n}_{\boldsymbol{B}}) \tag{12}$$

Ce résultat est démontré en annexe A.

La posterior marginalisée sur le binaire q_k est donc :

$$\begin{aligned}
[q_k|\infty_{-q_k,-\boldsymbol{c}}] &\propto \frac{\mathrm{e}^{-\sum_i \boldsymbol{y}_i^H \boldsymbol{B}(q_k)^{-1}\boldsymbol{y}_i}}{|\boldsymbol{B}(q_k)|^{I_s}} l^{q_k}(1-l)^{1-q_k} \\
&\propto \mathrm{e}^{-\operatorname{Trace}\left(\boldsymbol{B}(q_k)^{-1}\boldsymbol{S}_{yy}\right) - I_s|\boldsymbol{B}(q_k)| - q_k \ln\left(\frac{1}{l}-1\right)} \\
&\propto \mathrm{e}^{-g(q_k)}
\end{aligned}$$

De la même manière, la probabilité que le binaire change d'état est :

$$\begin{aligned}
\left[q_{k_{mod}}|\infty_{-q_{k_{mod}},-\boldsymbol{c}}\right] &\propto \frac{\mathrm{e}^{-\sum_i \boldsymbol{y}_i^H \boldsymbol{B}(q_{k_{mod}})^{-1}\boldsymbol{y}_i}}{|\boldsymbol{B}(q_{k_{mod}})|^{I_s}} l^{q_{k_{mod}}}(1-l)^{1-q_{k_{mod}}} \\
&\propto \mathrm{e}^{-g(q_{k_{mod}})}
\end{aligned}$$

Après normalisation, le changement d'état du binaire est accepté si

$$-\ln\left(\frac{1}{t} - 1\right) < g(q_k) - g(q_{k_{mod}})$$

.

4.1. Paramètres de la simulation

- $M = 20$
- $I_s = 10^4$
- SNR = 0 et -10 dB
- Nombre de fonction de base : 5
- Nombre de facteurs recherché : $M - 1$
- Trace($\boldsymbol{S}_{pp}$)/M = 2.4

4.2. Sans marginalisation

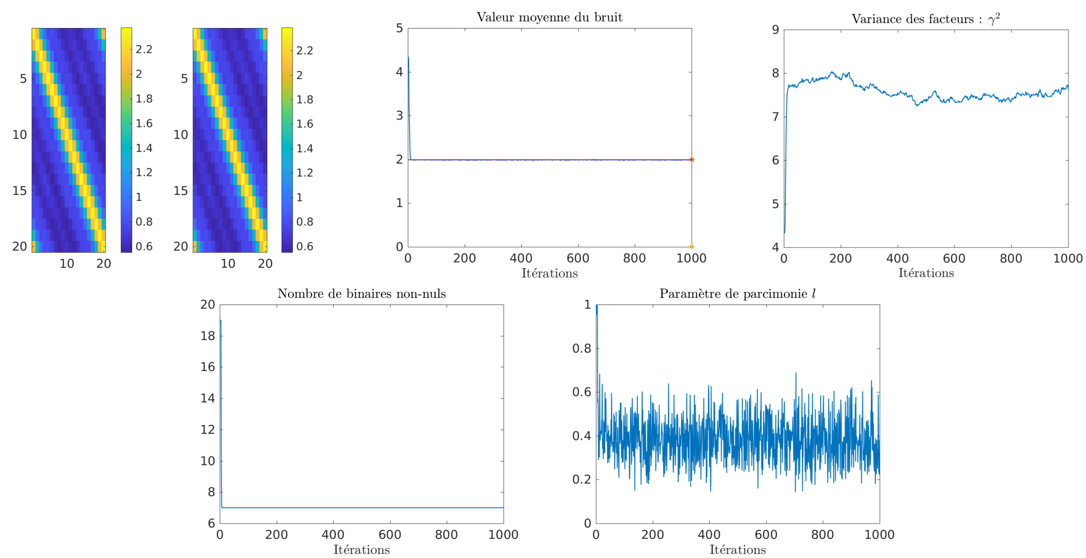

FIGURE 1 – *SNR = 0dB*

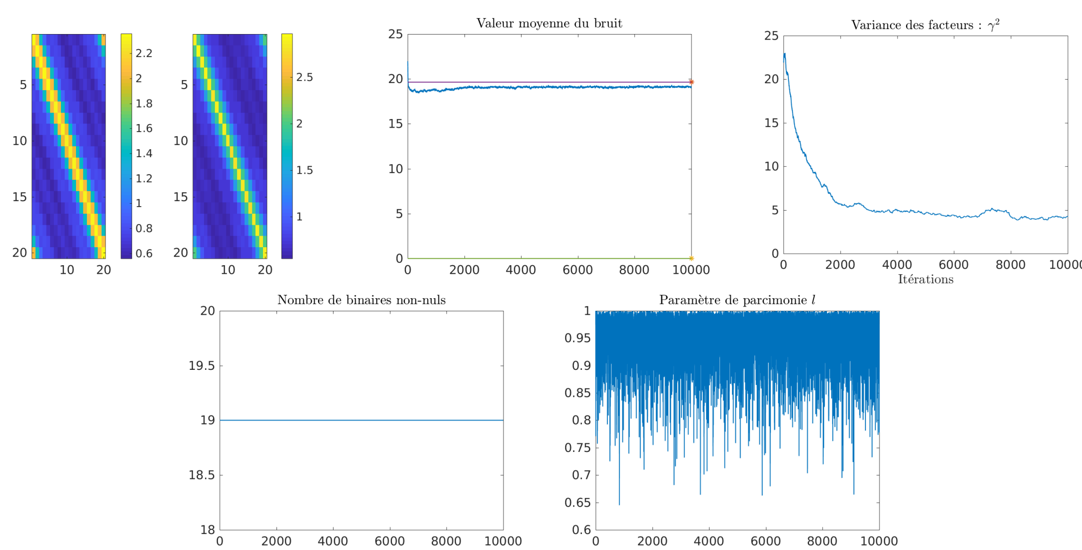

FIGURE 2 – *SNR = -10dB*

4.3. Avec marginalisation

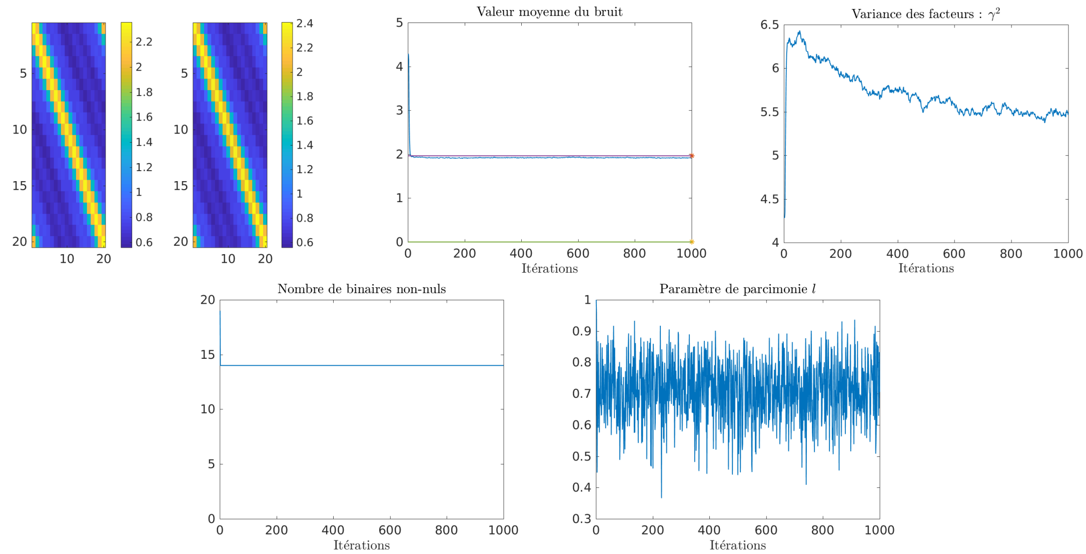

FIGURE 3 – *SNR = 0dB*

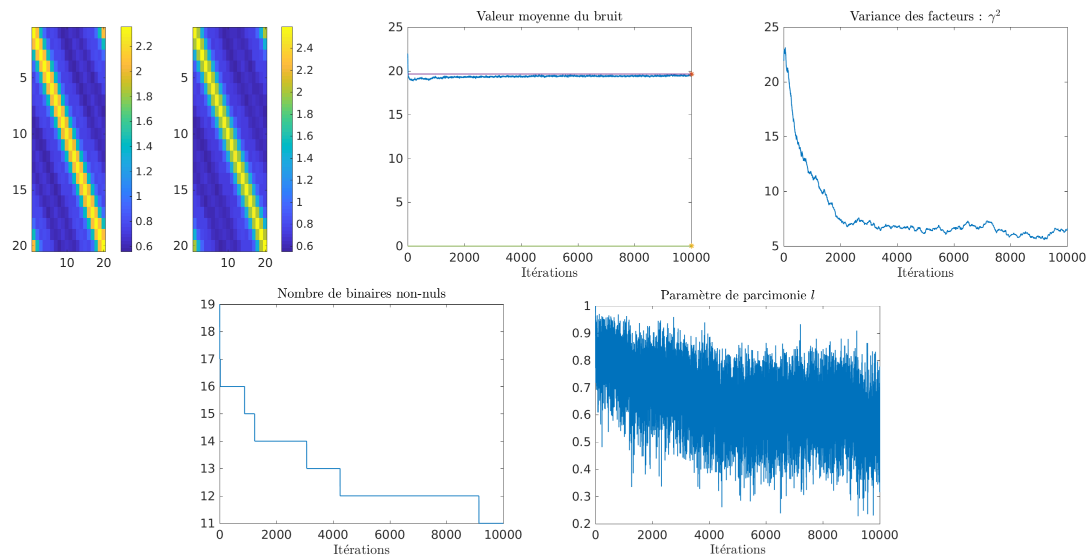

FIGURE 4 – *SNR = -10dB*

4.4. Comparaison avec l'ancienne version

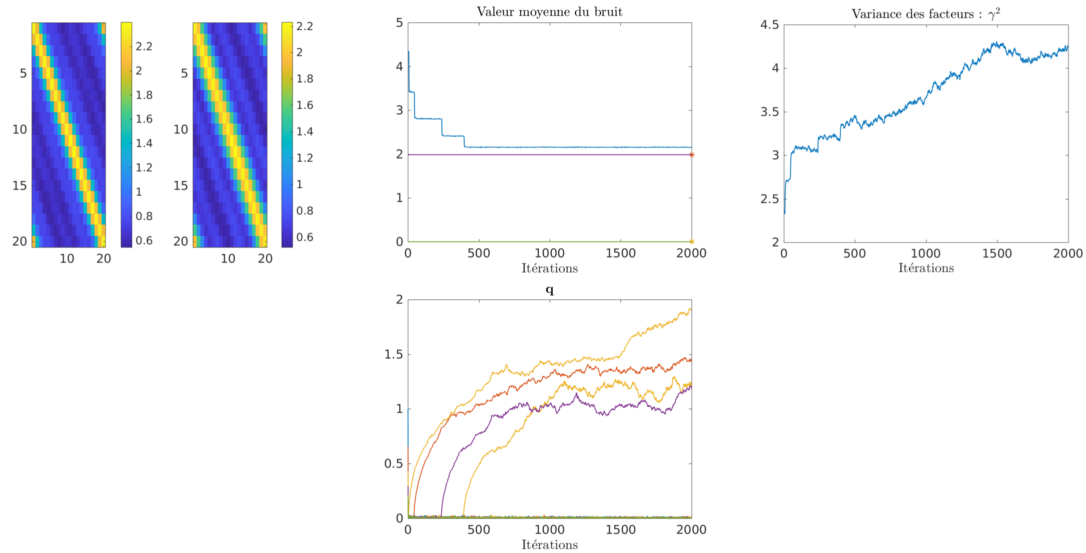

FIGURE 5 – *SNR = 0dB*

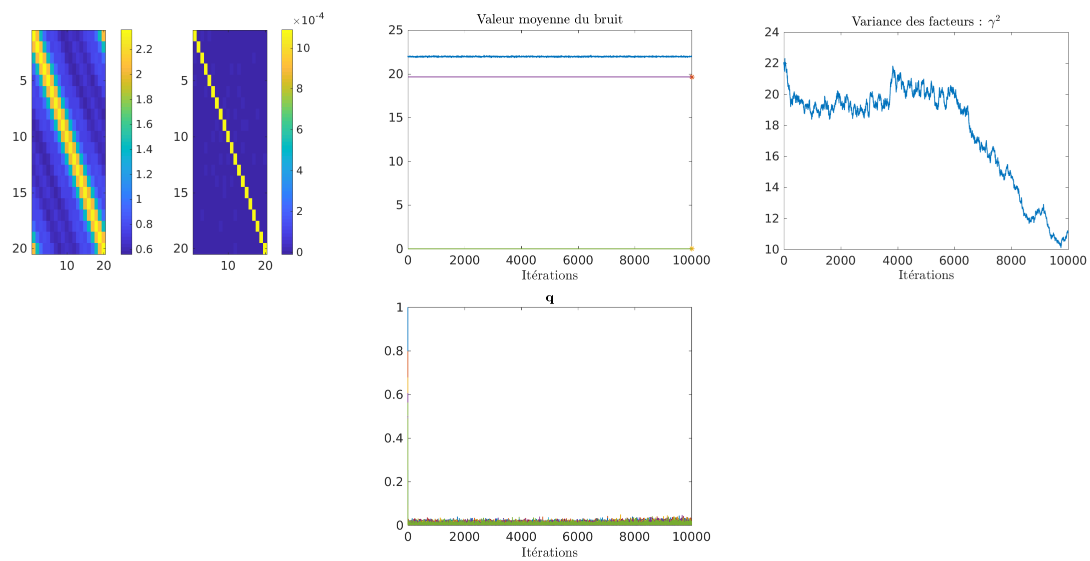

FIGURE 6 – *SNR = -10dB*

Remarques :

- Les résultats issus du modèle 2 changent beaucoup en fonction de l'initialisation, mais comme la convergence est rapide (surtout quand le SNR n'est pas trop bas), il serait possible de lancer simultanément plusieurs chaîne avec différentes initialisations, plus courtes.
- Il serait intéressant de comparer les résultats dans le cas où le nombre de fonction de bases est grand (hypothèse de parcimonie non respectée)

Annexe A. Détail du calcul de la postérieure marginalisée

La marginalisation revient à projeter la loi conjointe $[\boldsymbol{y}_i, \boldsymbol{c}_i]$sur $\boldsymbol{c}_i$:

$$
\begin{aligned}
[\boldsymbol{y}_i|\infty_{-\boldsymbol{c}}] &= \int [\boldsymbol{y}_i, \boldsymbol{c}_i]\,\mathrm{d}\boldsymbol{c}_i \\
&= \int [\boldsymbol{y}_i|\boldsymbol{c}_i][\boldsymbol{c}_i]\,\mathrm{d}\boldsymbol{c}_i
\end{aligned}
$$

Or,

$$
\begin{aligned}
[\boldsymbol{y}_i|\boldsymbol{c}_i][\boldsymbol{c}_i] &= \mathrm{e}^{(\boldsymbol{y}_i-\boldsymbol{Lqc}_i)^H\boldsymbol{\sigma_n^{-2}}(\boldsymbol{y}_i-\boldsymbol{Lqc}_i)}\mathrm{e}^{\boldsymbol{c}^h\boldsymbol{\gamma}^{-2}\boldsymbol{c}_i} \\
&= \mathrm{e}^{-\boldsymbol{y}_i^H\boldsymbol{\sigma}_n^{-2}\boldsymbol{y}_i}\ \mathrm{e}^{\underbrace{(\boldsymbol{Lqc}_i)^H\boldsymbol{\sigma}_n^{-2}\boldsymbol{y}_i}_{\boldsymbol{c}_i\boldsymbol{\Omega}_{\boldsymbol{c}_i}^{-1}\boldsymbol{\mu}_i}}\ \mathrm{e}^{\underbrace{\boldsymbol{y}_i^H\boldsymbol{\sigma}_n^{-2}\boldsymbol{Lqc}_i}_{\boldsymbol{\mu}_i^H\boldsymbol{\Omega}_{\boldsymbol{c}_i}^{-1}\boldsymbol{c}_i}}\ \mathrm{e}^{\underbrace{-(\boldsymbol{Lqc}_i)^H\boldsymbol{\sigma}_n^{-2}\boldsymbol{Lqc}_i}_{-\boldsymbol{c}_i^H\boldsymbol{\Omega}_{\boldsymbol{c}_i}^{-1}\boldsymbol{c}_i}} \\
&= \boldsymbol{C}_1\mathrm{e}^{-(\boldsymbol{c}_i-\boldsymbol{\mu}_i)^H\boldsymbol{\Omega}_{\boldsymbol{c}_i}(\boldsymbol{c}_i-\boldsymbol{\mu}_i)}
\end{aligned}
$$

avec $\boldsymbol{\mu}_i$ et $\boldsymbol{\Omega}_{\boldsymbol{c}_i}$ sont donnés dans l'encart (5). Donc,

$$
[\boldsymbol{y}_i|\infty_{-\boldsymbol{c}}] = \int [\boldsymbol{y}_i|\boldsymbol{c}_i][\boldsymbol{c}_i]\,\mathrm{d}\,\boldsymbol{c}_i = \boldsymbol{C}_1 \underbrace{\int \mathcal{N}_c(\boldsymbol{\mu}_i, \boldsymbol{\Omega}_{\boldsymbol{c}_i})\,\mathrm{d}\boldsymbol{c}_i}_{=1} = \boldsymbol{C}_1
$$

Par identification,

$$
\begin{aligned}
&\boldsymbol{C}_1\ \mathrm{e}^{-\boldsymbol{\mu}_i^H\boldsymbol{\Omega}_{\boldsymbol{c}_i}\boldsymbol{\mu}_i} = \mathrm{e}^{-\boldsymbol{y}_i^H\boldsymbol{\sigma}_n^{-2}\boldsymbol{y}_i} \\
\Leftrightarrow\quad &\boldsymbol{C}_1 = \mathrm{e}^{-\boldsymbol{y}_i^H\boldsymbol{\sigma}_n^{-2}\boldsymbol{y}_i}\ \mathrm{e}^{\underbrace{\boldsymbol{y}_i^H\boldsymbol{\sigma}_n^{-2}\boldsymbol{Lq\Omega}_{\boldsymbol{c}_i}\boldsymbol{q}^H\boldsymbol{L}^H\boldsymbol{\sigma}_n^{-2}\boldsymbol{y}_i}_{\boldsymbol{A}}} \\
&\boldsymbol{C}_1 = \mathrm{e}^{-\boldsymbol{y}_i^H\boldsymbol{B}^{-1}\boldsymbol{y}_i}
\end{aligned}
$$

avec

$$
\boldsymbol{A} = \frac{\boldsymbol{\sigma}_n^{-2}}{1 + \underbrace{\boldsymbol{L}^{H^{-1}}\boldsymbol{q}^{H^{-1}}\boldsymbol{\gamma}^{-2}\boldsymbol{q}^{-1}\boldsymbol{L}^{-1}}_{\boldsymbol{A}_2^{-1}}\boldsymbol{\sigma}_n^2}
$$

et

$$
\begin{aligned}
\boldsymbol{B}^{-1} = -\boldsymbol{A} + \boldsymbol{\sigma}_n^{-2} &= \frac{-\boldsymbol{\sigma}_n^{-2} + \boldsymbol{\sigma}_n^{-2}\left(1 + \boldsymbol{A}_2^{-1}\boldsymbol{\sigma}_2^2\right)}{1 + \boldsymbol{A}_2^{-1}\boldsymbol{\sigma}_n^2} \\
&= \frac{1}{\boldsymbol{A}_2 + \boldsymbol{\sigma}_n^2}
\end{aligned}
$$

Finalement,

$$
[\boldsymbol{y}_i|\infty_{-\boldsymbol{c}}] = \boldsymbol{C}_1 = \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{B}) \quad \text{où} \quad \boldsymbol{B} = \boldsymbol{Lq}\gamma^2\boldsymbol{q}^H\boldsymbol{L}^H + \boldsymbol{\sigma}_n^2.
$$

Références

Di Ge, Jérôme Idier, and Eric Le Carpentier. Enhanced sampling schemes for mcmc based blind bernoulli–gaussian deconvolution. *Signal Processing*, 91(4) :759–772, 2011.